

Draft

User Manual

SimSalmon version 4.5.2 beta

Paul McElhany

National Marine Fisheries Service
Northwest Fisheries Science Center

&

John Payne

University of Washington
Department of Zoology

July 24, 2001

Table of Contents

Table of Contents	2
General Disclaim	4
SimSalmon Overview	4
Modified Dennis Analysis Module	6
Modified Dennis Analysis Overview	6
Dennis-Holmes Parameters and Extinction Risk Sub-Module	8
Dennis-Holmes Parameters and Extinction Risk Input Window	10
Dennis-Holmes Parameters and Extinction Risk Output	16
Viability Criteria Overview	19
Viable Criteria Option of Dennis-Holmes Parameters and Extinction Risk Sub-module	21
Viable Criteria Test Sub-module	21
Extinction Risk Table Sub-Module	21
Minimum Size Table Sub-module	24
Simulate Trajectories Module	27
Simulate Trajectories Overview	27
Density-independent population growth: the Dennis Model	28
Density-dependent growth	28
Density-Dependent Models	28
Parameters that control "error" added to the simulations	29
How to use the <i>Simulate Trajectories</i> module	30
Bottom Pane Parameters	32
Base Output File	33
Evaluate Criteria Module	35
Evaluate Criteria Overview	35
How to use the <i>Evaluate Criteria</i> module	35
Output	37
Recruits per Spawner Module	39
Recruits per Spawner Overview	39
Calculate Spawners/Spawner Sub-Module	41
Input Window	42
Output	42
Extinction Calculator Sub-Module	44
Input Windows	44
Extinction Calculator Single Mode Output File:	46
Batch Mode:	46
Sensitivity Analysis Sub-Module	48
Patterns in Spawners/Spawner Sub-Module	49
Temporal Autocorrelation	49
Trends in Spawners per Spawner	50
Appendix A: Correction for hatcheries and harvest	52
Slope Method of Parameter Estimation Corrections for Hatchery	52
Slope Method of Parameter Estimation Corrections for Harvest	56
Slope Method of Parameter Estimation Corrections for Hatchery and Harvest	60
Appendix B: Variance degrees of freedom	61

General Disclaim

This is a beta version of this program, meaning it is still in development and testing. The program may contain errors that affect output. Please report any suspected bugs to Paul McElhany (paul.mcelhany@noaa.gov).

The program uses a number of different algorithms for estimating parameters related to the extinction risk for salmon populations. All approaches for analyzing extinction risk have severe limitations and simplifying assumptions. The models in SimSalmon are no exception and users of the program must have an understanding of these limitations before attempting to interpret the program results.

SimSalmon Overview

SimSalmon is a collection of computers modules for conducting viability analyses for Pacific salmonids based on time series of abundance and other data. There are four major modules (Figure 1) 1) Modified Dennis, 2) Recruits per Spawner, 3) Simulate Trajectories, 4) Evaluate Criteria. The analyses performed by each of the modules and sub-modules are shown in Table 1. The program was developed under the direction of Paul McElhany (NMFS-NWFSC) and was written in Java by Paul McElhany, Frank Neo (Aris Corp.) and John Payne (University of Washington). The program was developed at the Northwest Fisheries Science Center. SimSalmon may be freely distributed, but the general disclaimer must be included with distribution.

Table 1: Modules and sub-modules of SimSalmon.

Module	Sub-module	Description
Modified Dennis	Dennis-Holmes parameters and extinction risk	From a user specified time series, this module fits parameters for a stochastic exponential growth model, which is used to estimate extinction risk.
	Viable Criteria Test	Evaluates whether or not a time series exceeds user modified viability criteria.
	Extinction Risk Table	Uses Dennis model to estimate extinction risk as a function of mean growth rate and process variance.
	Minimum Size Table	Estimates minimum viable population size as a function of mean growth rate and process variance.
Recruits per Spawner Analysis	Calculate Spawners/Spawner	Calculates recruits (as post harvest spawners) per spawner and provides simple summary statistics.
	Extinction Probability Calculator	Uses a partially age structured simulation model and a time series derived distribution of recruits per spawner to estimate extinction risk.
	Sensitivity Analysis	Allows systematic exploration of various parameters on the extinction risk estimates using the recruits per spawner model.
	Patterns in Spawners/Spawner	Explores autocorrelation and temporal trend in recruits per spawner extinction model.
Simulate Trajectories	Simulate Trajectories	Simulates partially age structured population with user specified parameters. Output is in form read by the Modified Dennis and Recruits per Spawner modules
	Trajectory Summary	Provides summary statistics from a collection of simulated trajectories.
	Subsample Trajectories	A utility that creates new trajectories by subsampling input trajectories.
Evaluate Criteria		Evaluates the Modified Dennis viability criteria by examining the fraction of type I and type II errors made when analyzing simulated trajectories.

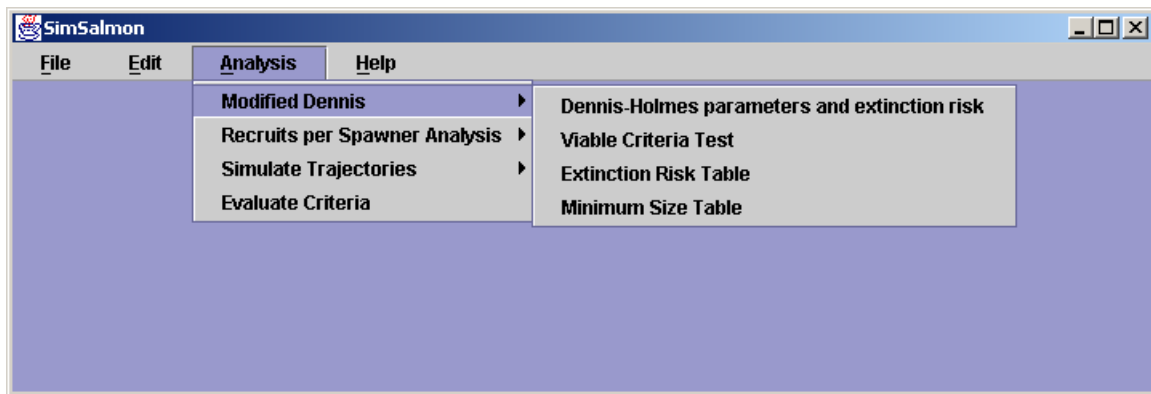


Figure 1: Main analysis modules of SimSalmon and the sub-modules of the Modified Dennis analysis.

Modified Dennis Analysis Module

Modified Dennis Analysis Overview

The Dennis et. al (1991) model is based on a simple stochastic exponential growth model,

$$N_{t+1} = N_t e^{\mu + \sigma Z_t}, \text{ eq. 1}$$

where N_t is the population size at time t , μ is rate of increase or decrease in the population, σ^2 is environmentally induced variance and Z_t is a standard normal deviate. This model describes a diffusion process with drift. The extinction risk for a population can be estimated as the probability that the population declines from some initial size to some lower size within a specified amount of time. For projecting into the future, the initial size is typically set as the current population size. The lower size is referred to as the quasi-extinction threshold (QET) and is the value at which the population is considered functionally extinct. The equation for this extinction probability is given as

$$\Pr(N_0 \rightarrow N_q \text{ within } t_e) = \Phi\left(\frac{-\ln\left(\frac{N_0}{N_q}\right) + |\mu|t_e}{\sigma\sqrt{t_e}}\right) + e^{\left(\frac{2\ln\left(\frac{N_0}{N_q}\right)|\mu|}{\sigma^2}\right)} \Phi\left(\frac{-\ln\left(\frac{N_0}{N_q}\right) - |\mu|t_e}{\sigma\sqrt{t_e}}\right), \text{ eq. 2}$$

where N_0 is the initial population size, N_q is the QET, t_e is the time horizon, Φ is the cumulative normal distribution function, μ is the mean of the process and σ^2 is the process variance. In Dennis et al, the parameters μ and σ^2 are basically estimated as

$$\mu = \text{mean}\left(\ln\left(\frac{N_{t+1}}{N_t}\right)\right), \text{ eq. 3,}$$

$$\sigma^2 = \text{var}\left(\ln\left(\frac{N_{t+1}}{N_t}\right)\right), \text{ eq. 4}$$

where N_t is the population count at time t . In this model, the ratio $\ln(N_{t+1}/N_t)$ is approximately normally distributed.

SimSalmon implements a parameter estimation technique developed by Holmes (2001) to correct for the tremendous biases in parameter estimation that can occur in applying the Dennis method to datasets that sample a single life stage or that contain large measurement error. Salmon counts are typically for a single life stage (spawners) and typically contain large measurement errors. The Holmes (2001) parameter estimation method corrects for bias in two different ways. First is the use of a running sum to correct for the measurement of only a single life stage (spawners). The running sum may either be a weighted running sum that considers the age structure of the spawners to approximate the total number of current and future spawners or an unweighted sum that simply adds up the number of spawners over a fixed number of years. As the method is

not particularly sensitive to age structure, the results from either approach are often similar (Holmes 2001). The weighted sum is proportional to the total population size, N_t , and is estimated by the weighted running sum

$$N_t = S_t + \sum_{i=1}^{\max \text{ age}-1} w_i S_{t-i}, \quad \text{eq. 5.}$$

where S_t is the number of spawners observed in year t and

$$w_i = 1 - \sum_{j=1}^i \text{fractionAge}_j, \quad \text{eq. 6}$$

where fractionAge_j is the average fraction of spawners of age j .

The second modification to parameter estimation helps to correct for the upward bias in the σ^2 estimate created by measurement error. The approach, referred to as the “slope method” (Holmes 2001), uses the slope of $\sigma^2(\tau)$ vs. τ as an estimate of σ^2 , where

$$\sigma^2(\tau) = \text{var} \left(\ln \left(\frac{N_{t+\tau}}{N_t} \right) \right), \quad \text{eq. 7}$$

The value τ is referred to as the lag. Details of this method are provided in Holmes (2001). Basically, the approach works because process error (which we want to estimate) contributes to the increase in the variance estimate with increasing lag, whereas measurement error does not. Under the assumptions of the model, the relationship of $\sigma^2(\tau)$ vs. τ should be linear, and this is an important assumption to examine. A common cause of non-linearity in the relationship is application to a time series that is too short.

The parameters that we need to estimate for calculating extinction risk and viability criteria are μ and σ^2 of the underlying process. Hatcheries and harvest can obfuscate the underlying process and SimSalmon includes several options to “correct” for harvest and hatchery effects on the time series. Appendix A describes the mathematical detail behind these correction methods. Ignoring the complications imposed by age structure, the corrections can be described as a modification of the N_{t+1}/N_t ratio. Instead of N_{t+1}/N_t , the ratio of harvest and/or hatchery activity is given in Table 2.

In SimSalmon, it is possible to correct for both harvest and hatchery simultaneously, either one individually or neither to explore the consequences of different assumptions.

The extinction risk estimation uses equation 2. The initial population size, N_0 , is taken as the current (or last) running sum value in the time series. User input controls the time horizon, the quasi-extinction threshold, confidence level for confidence intervals, the number of lags for the slope estimate of the variance, and the method of calculating the running sum.

Table 2: Modifications to the N_{t+1}/N_t ratio to correct for harvest and hatchery impacts on the time series. The total number of spawners is S_t , W_t is the number of natural origin spawners at time t , C_t is the additional number of natural origin fish that would have returned to spawn if there had been no harvest, H_t is the number of hatchery origin fish that spawn in the wild and δ is the reproductive success of hatchery origin fish spawning in the wild relative to natural origin fish. These equations ignore the complications of age structure, which are dealt with in Appendix A.

Correct for harvest	Correct for hatchery	Ratio
---------------------	----------------------	-------

No	No	$\frac{S_{t+1}}{S_t}$
No	Yes	$\frac{W_{t+1}}{W_t + \delta H_t}$
Yes	No	$\frac{S_{t+1} + C_{t+1}}{S_t}$
Yes	Yes	$\frac{W_{t+1} + C_{t+1}}{W_t + \delta H_t}$

Dennis-Holmes Parameters and Extinction Risk Sub-Module

The Dennis-Holmes Parameters and Extinction Risk sub-module is the primary tool for evaluating the status of a population. The input window for this sub-module is shown in Figure 2. The sub-module takes as input a time series of abundance and other data. The sub-module can return a number of different outputs depending on user specified options. The main potential outputs are the population growth rate, the population extinction risk and critical viability criteria.

SimSalmon

File Edit Analysis Help

Dennis-Holmes parameters and extinction risk

Input Data

Add file/s to analysis... Remove selected file

☐ Input Data includes Harvest column

☐ Replace missing values with average of adjacent years

Begin analysis with year 1 and end with year 3,000

Output File

Output File Path Browse...

Parameter Estimation

☐ Correct for Hatchery Spawners Effectiveness of Hatchery Spawners 0.5

☐ Correct for Harvest

☐ Use unweighted running sum of length 4

Maximum Tau 4

Extinction Risk

Quasi-Extinction Threshold: 1

Maximum Time Horizon 100 PE Report Times 25 50 100

Confidence Interval

Parameter and Extinction Risk Confidence Level (%) 95

Viability Criteria

☐ Calculate Viability Criteria The acceptable risk for the viability criteria is a 5 percent chance

of declining to a running sum of 1 fish in 100 years. Number of Replicates for simulation: 1,000

Growth rate confidence level (%) for min. size criteria: 95

Minimum size analysis range: 11 to 100,000

Run!

Figure 2: Dennis-Holmes analysis input window for SimSalmon v 4.5.3.

DENNIS-HOLMES PARAMETERS AND EXTINCTION RISK INPUT WINDOW

Input Data

Add files to batch analysis...

Selecting the “Add file/s to batch analysis...” button brings up the file chooser dialog box. Navigate through the chooser to select all the input text files. There are two types of input files, which differ only in whether or not data on harvest are included in the file. Input files are tab-delimited text files. The first row of the file contains column headings. Column headings can contain any characters and numbers, can be of any length, but should not contain spaces. For input files without harvest data, the first column should be “year”; the second “abundance”, the third “fraction_of_natural_origin_spawnners” and the next columns should be “Age1” to “AgeX”, where X is the maximum age at reproduction. Columns headings do not have to have these names, but the data columns must be arranged in this order. The “year” is the year of the observation. Years should be entered sequentially without any gaps in the numbering of years. “Abundance” is the total number of spawners observed, regardless of origin. The “fraction_of_natural_origin_spawnners” is a value between zero and one that is the fraction of natural origin spawners. Natural origin spawners are spawners whose parents spawned in the wild. The “fraction_of_natural_origin_spawnners” is equal to one minus the fraction of first generation hatchery fish spawning in a given year. If all hatchery fish are marked, the “fraction_of_natural_origin_spawnners” is the fraction of unmarked fish observed in a given year. The “Age_” columns should contain the fraction of individuals of each age class that were observed spawning in a given year. Note that the age proportions refer to the observations within a given year, not to the distributions of recruits from that year. The “Age_” values in a single row should sum to one. If only a single or average estimate of age structure is available, the same values can be put in the age structure columns for every year. Holmes (2001) has shown that the parameter estimation method is not very sensitive to errors in the age structure and some options in the analysis do not use the annual age data at all. If data are missing for any year, the missing value “-99” should be placed in the abundance column. Table 3 is an example of this type of input data file.

Table 3: Example input data file with no harvest column.

year	Spawners	fracWild	age1	age2	age3	age4	age5	age6
1977	774	0.818	0	0	0.024	0.534	0.292	0.150
1978	694	0.430	0	0	0.031	0.517	0.410	0.041
1979	371	0.713	0	0	0.042	0.533	0.373	0.051
1980	1025	0.761	0	0	0.011	0.679	0.281	0.029
1981	2150	0.776	0	0	0.000	0.449	0.481	0.070
1982	869	0.784	0	0	0.007	0.432	0.507	0.055
1983	532	0.609	0	0	0.055	0.268	0.614	0.063
1984	943	0.470	0	0	0.061	0.614	0.273	0.053
1985	632	0.592	0	0	0.019	0.505	0.467	0.009
1986	919	0.363	0	0	0.044	0.569	0.331	0.056
1987	982	0.547	0	0	0.039	0.419	0.477	0.065
1988	1079	0.505	0	0	0.008	0.569	0.393	0.029
1989	500	0.650	0	0	0.033	0.622	0.316	0.029
1990	-99	-99	-99	-99	-99	-99	-99	-99
1991	959	0.744	0	0	0.040	0.429	0.484	0.048
1992	1974	0.693	0	0	0.025	0.651	0.286	0.038
1993	843	0.730	0	0	0.050	0.323	0.552	0.076
1994	725	0.792	0	0	0.033	0.726	0.202	0.039
1995	1030	0.783	0	0	0.027	0.562	0.375	0.036
1996	725	0.452	0	0	0.020	0.628	0.329	0.024
1997	456	0.900	0	0	0.047	0.602	0.333	0.018

The second type of input data file is the same as described above, except for the inclusion of a “harvest” column (Table 4). The “harvest” column goes between the “fraction_of_natural_origin” and the “Age1” column. The harvest column is the number of additional natural origin spawners that would have been observed in a given year if there had been no harvest. This column does NOT contain the number of fish harvested in a given year. This is a difficult parameter to estimate and requires estimates and assumptions about harvest rates and age structure. As every system is likely to present unique quirks in estimating this parameter, we are not providing any general guidance on estimation methods. If the harvest analysis is used, the age structure depicted in the “Age_” columns should reflect the age structure that would have been observed if there had been no harvest. As noted above, the Dennis-Holmes method is not overly sensitive to errors in age structure.

Table 4: Example input data file with harvest column.

year	Spawners	fracWild	harvest	age1	age2	age3	age4	age5	age6
1977	774	0.818	1097	0	0	0.024	0.534	0.292	0.150
1978	694	0.430	832	0	0	0.031	0.517	0.410	0.041
1979	371	0.713	844	0	0	0.042	0.533	0.373	0.051
1980	1025	0.761	2978	0	0	0.011	0.679	0.281	0.029
1981	2150	0.776	1075	0	0	0.000	0.449	0.481	0.070
1982	869	0.784	1621	0	0	0.007	0.432	0.507	0.055
1983	532	0.609	738	0	0	0.055	0.268	0.614	0.063
1984	943	0.470	854	0	0	0.061	0.614	0.273	0.053
1985	632	0.592	799	0	0	0.019	0.505	0.467	0.009
1986	919	0.363	148	0	0	0.044	0.569	0.331	0.056
1987	982	0.547	217	0	0	0.039	0.419	0.477	0.065
1988	1079	0.505	90	0	0	0.008	0.569	0.393	0.029
1989	500	0.650	74	0	0	0.033	0.622	0.316	0.029
1990	-99	-99	-99	-99	-99	-99	-99	-99	-99
1991	959	0.744	5	0	0	0.040	0.429	0.484	0.048
1992	1974	0.693	204	0	0	0.025	0.651	0.286	0.038
1993	843	0.730	72	0	0	0.050	0.323	0.552	0.076
1994	725	0.792	9	0	0	0.033	0.726	0.202	0.039
1995	1030	0.783	15	0	0	0.027	0.562	0.375	0.036
1996	725	0.452	38	0	0	0.020	0.628	0.329	0.024
1997	456	0.900	2	0	0	0.047	0.602	0.333	0.018

Input data includes harvest column

The input data includes harvest column checkbox is used to indicate which type of data set is being used. All the data files in a single run must be of the same type. Having a harvest column in the data set does not require that analyses correct for harvest, as correcting for harvest is a separate user defined option.

Replace missing values with average of missing years

If this option is selected, simsalmon will replace missing data values with the arithmetic average of data from the adjacent years. The algorithm averages the values in the year immediately previous to the missing year(s) and immediately after the missing year(s). This option is often the only way to get a time series long enough for any analysis.

Begin analysis with year ___ and end with year ___

These parameters can be used to analyze a subset of years in the data set. The program will analyze years greater than or equal to the first value and less than or equal to the second value. To analyze the entire data set, simply accept the defaults. The defaults cause the program to analyze years 1 to 3000 and the actual years of data are sure to fall in this range. To analyze all years from 1980 on, enter 1980 for the first value and 3000 for the second. To analyze values from 1980 to 1995, enter 1980 for the first value and 1995 for the second.

Parameter Estimation***Correct for hatchery spawners***

The correct for harvest check box allows the user to estimate what growth rate and extinction risk would have been (and potentially could be in the future) if there is no hatchery input of spawners. Ignoring the complications of age structure, the modifications to the N_{t+1}/N_t ratio that correct for harvest are shown in Table 1. The equations that include age structure are in Appendix A. The correction for hatchery spawners is merely a demographic accounting of hatchery input and does not consider any ecological interactions between hatchery and natural origin fish.

Effectiveness of hatchery spawners

This is the reproductive success of hatchery origin fish spawning in the wild relative to natural origin fish spawning in the wild. Natural origin spawners are spawners whose parents spawned in the wild and hatchery origin spawners are spawners whose parents were spawned in the hatchery. The effectiveness of hatchery spawners is the δ parameter in Table 1. Values will be non-negative. A value of zero indicates that hatchery origin fish to not reproduce successfully in the wild. A value of one indicates that hatchery origin fish have a reproductive success in the wild equal to that of natural origin fish. In practice, values for this parameter will typically be greater than zero and less than one. However, values greater than one are possible if hatchery origin fish have a higher reproductive success than natural origin fish in the wild. This input field is disabled unless the “Correct for hatchery spawners” checkbox is selected.

Correct for harvest

The correct for harvest check box allows the user to estimate what growth rate and extinction risk would have been (and potentially could be in the future) if there is no harvest. The check box is disabled unless the data set has a harvest column and the “Input data includes harvest column” check box is selected. Ignoring the complications of age structure, the modifications to the N_{t+1}/N_t ratio that correct for harvest are shown in Table 1. The equations that include age structure are in Appendix A.

Use unweighted running sum of length

The default running sum calculation uses the weighting in equation 5. Selecting the “Use unweighted running sum of length” checkbox changes the running sum to an unweighted running sum of the user specified length. The unweighted running sum is the method used by Holmes and Fagan (unpublished manuscript) and since the Dennis-Holmes method is not overly sensitive to the age structure or weighting function, results from either type of running sum tend to be similar (though not always). SimSalmon uses slightly different equations for the two types of running sums when estimating the variance degrees of freedom for determining confidence intervals. (See appendix B). Because of the variance degrees of freedom equations, only the values of 3, 4 or 5 are allowed as input for the running sum length. Holmes and Fagan (unpublished manuscript) have shown that values in this range provide the most accurate estimates of population growth rate.

Maximum Tau

The process variance is estimated as the slope of $\sigma^2(\tau)$ vs. τ . The maximum tau value is the maximum number of lags used to estimate the variance. Work by Holmes (2001) indicates values around four provide the best estimation of the process variance. However, if the time series is very short (e.g. <15-20 yrs), the $\sigma^2(\tau)$ vs. τ relationship tends to violate the linearity assumption as larger lag distances show lower variance than expected. This can occur because the sample size for the variance estimate declines with larger tau and the variance estimate tends to decline with shorter sample size. One possible solution to improve the σ^2 estimate is to decrease the maximum tau value to three or even two. The Maximum Tau values must be integers greater than or equal to one. If a value of one is entered, the variance estimate does not use slope method and σ^2 is calculated simply as $\text{var}(\ln(N_{t+1}/N_t))$.

Extinction Risk

Quasi-Extinction threshold

This is the quasi-extinction threshold (QET) used for the extinction risk estimation (i.e. N_q in equation 2). The quasi-extinction threshold is a value at which the population is considered functionally extinct. If the population ever reaches this value at any time between year one and the time horizon the population is considered extinct. The quasi-extinction value is set based on the running sum, not the number of spawners. The relationship between the running sum and the number of spawners depends on the user selected running sum options (see below) and, depending on options selected, the population age structure. A QET of one generally indicates true extinction, because it represents only one fish over the entire running sum. The extinction risk model can be

very sensitive to the QET value. Unfortunately, it is extremely difficult to empirically determine an appropriate QET value.

Maximum time horizon

This parameter is the maximum number of years for which the program will calculate the extinction risk. Extinction risk, estimated using equation 2, is calculated for every year into the future from year one to the maximum time horizon. If the extinction risk in the next 30 years is all that is of interest, the input value is 30. If the time horizon of interest is the next 500 years, the input value is 500. Values must be integers greater than one.

PE report times

The PE report times are the time horizons for the extinction probabilities that are reported in the analysis summary file. The program actually calculates the extinction risk for all time horizons between one and “Maximum time horizon” and these values are output in the detailed population output file (see output descriptions below). However, only three time horizons, the PE report times, are output in the analysis summary file. The PE report time values must be between one and “Maximum time horizon”.

Confidence Interval

Parameter and extinction risk confidence level

Confidence levels are typically 90 or 95 percent, but can be set lower or higher. Confidence intervals for μ and σ^2 are estimated based on the approximate normal distribution of $\ln(N_{t+1}/N_t)$. Holmes and Fagan (unpublished manuscript) explore the estimation of confidence intervals when using the slope method. The confidence interval for μ is approximated as

$$\mu_{C.I.} \approx \hat{\mu} \pm Tinv(1 - \alpha / 2, df) * \frac{\sigma}{\sqrt{n}}, \text{ eq. 8}$$

where $Tinv(p, df)$ is the inverse student's t distribution function, α is the confidence level/100, df is the degrees of freedom for the variance distribution and n is the number of $\ln(N_{t+1}/N_t)$ values calculated from the time series. Because SimSalmon uses the slope method to estimate the variance, the degrees of freedom for the variance estimate is not simply the number of $\ln(N_{t+1}/N_t)$ values minus one. The slope method reduces the bias of the variance estimate when data contain measurement error, but at a cost of decreased precision. Holmes and Fagan (unpublished manuscript) provide an approximation for the degrees of freedom as a function of the length of the time series and the length of an unweighted running sum (see below about running sum). SimSalmon uses an approximation of the degrees of freedom base on, but not identical to, the tables in Holmes and Fagan. Details of the degrees of freedom approximation are in Appendix B.

The confidence interval for σ^2 is approximated as

$$\sigma_{LowerC.I.}^2 = \frac{df * \hat{\sigma}^2}{Xinv(\alpha / 2, df)}, \text{ eq. 9}$$

$$\sigma_{UpperC.I.}^2 = \frac{df * \hat{\sigma}^2}{Xinv((1 - \alpha / 2), df)}$$

where $Xinv(p, df)$ is the inverse chi² distribution and other parameters are as described above.

Confidence intervals for the extinction probabilities are estimated using a parametric bootstrap. Assuming the current population size is correctly estimated, the extinction risk for a given QET and time horizon depends only on μ and σ^2 . The parameters μ and σ^2 are estimated from the data and the approximate distribution of the parameters is known, with

$$\mu \approx \hat{\mu} - StudentT(df) * \frac{\sigma}{\sqrt{n}}, \text{ eq. 10}$$

$$\sigma^2 \approx \frac{df * \hat{\sigma}^2}{X(df)}$$

where $StudentT(df)$ is the Student's t distribution and $X(df)$ is the chi² distribution.

The distribution of the extinction probabilities is calculated by repeatedly taking independent random draws from the distributions of μ and σ^2 to obtain random μ , σ^2 pairs. The extinction risk is calculated for each random μ , σ^2 pair using the same initial population size, QET and time horizon. The results provide an approximation of the distribution of extinction risk. SimSalmon generates a distribution of 2,000 extinction risk estimates and the confidence limits are reported as the confidence level/2 percent tails of the distribution. Note that this provides an estimate of the extinction probability distribution based on the uncertainty associated with using a finite time series. It does not include the possibility that the model might be wrong...

Viability Criteria

The viability criteria approach and input parameters are Described in the section on the Viability Criteria sub-module.

DENNIS-HOLMES PARAMETERS AND EXTINCTION RISK OUTPUT

The Dennis-Holmes Parameters and Extinction Risk module produces two summary files that includes information on all the input files, plus two files for each of the input files. If the “viability criteria” option is selected, the program outputs a third file per input file. All files are tab-delimited text and it is convenient to import them into Excel. File names and descriptions are shown in Table 5.

Table 5: Output files from the “Dennis-Holmes Parameters and Extinction Risk” module. The *Outfilename* is the output file name specified by the user. The *inputfilename* is the name of an individual population input file.

Output File Name	Description
<i>Outfilename</i>	Summary file that contains population growth rate and extinction risk parameters.
<i>Outfilename_s2VsTau</i>	Summary file for plotting $\sigma^2(\tau)$ vs. τ .
<i>Outfilename_inputfilename_MDout</i>	General output with detail on parameter estimation and extinction risk for a specific population.
<i>Outfilename_inputfilename_NN</i>	Population specific data on $\ln(N_{t+1}/N_t)$ vs. time.
<i>Outfilename_inputfilename_crit</i>	Population specific viability criteria output.

Summary File

The summary file contains a header row, plus one row per input file. The column descriptions are shown in Table 6.

Table 6: Modified Dennis Summary file column descriptions. The values of @, #, ##, and ### are specified by the user.

Header Label	Description
inputFile	The name of the input file.
mu	$\text{Mean}(\ln(N_{t+1}/N_t))$
sigma2	$\text{Variance}(\ln(N_{t+1}/N_t))$, calculated using slope method.
alpha	Median annual population growth rate = e^μ
lambda	Average annual population growth rate = $e^{\mu + \frac{\sigma^2}{2}}$
geoMeanRunSum	Geometric mean of running sum
geoMeanSpawners	Geometric mean of total spawners
slopeLnSpawners	Slope of regression of $\ln(\text{total spawners})$ Vs. year
geoMeanNatSpawners	Geometric mean of natural origin spawners
slopeLnNatSpawners	Slope of regression of $\ln(\text{natural origin spawners})$ Vs. year
PE#	Probability of extinction in # years calculated using Dennis-Holmes method
PE##	Probability of extinction in ## years calculated using Dennis-Holmes method
PE###	Probability of extinction in ### years calculated using Dennis-Holmes method
Mu@CI	The @% confidence interval is $\mu \pm \text{Mu@CI}$
s2df	The degrees of freedom for the variance estimate
s2Lower@CI	Lower @% confidence interval on σ^2
s2Upper@CI	Upper @% confidence interval on σ^2
alphaLower@CI	Lower @% confidence interval on α .
alphaUpper@CI	Upper @% confidence interval on α .
PE#lower	Lower @% confidence interval on # year extinction probability.
PE#upper	Upper @% confidence interval on # year extinction probability.
PE##lower	Lower @% confidence interval on ## year extinction probability.
PE##upper	Upper @% confidence interval on ## year extinction probability.
PE###lower	Lower @% confidence interval on ### year extinction probability.
PE###upper	Upper @% confidence interval on ### year extinction probability.

Outfilename_s2VsTau File

The *Outfilename_s2VsTau* file can be used to plot σ^2 vs. τ . The first column is τ , which is followed by a column of $\sigma^2(\tau)$ estimates for each population. As described above, if the assumptions of the model hold, the relationship between σ^2 and τ should be nearly linear.

outfilename_inputfilename_Mdout File

The outputfile named *outfilename_inputfilename_Mdout* contains population specific output, with different sections arranged vertically in the file. The first section describes some of the user specified inputs. The second section provides some output summary as shown in Table 7.

Table 7: Output summary section from *outfilename_inputfilename_Mdout* file.

Output label	Description
Mean age at reproduction:	Average age at reproduction calculated by pooling all fish in all years
Mu:	Mean($\ln(N_{t+1}/N_t)$)
sigma2:	Variance($\ln(N_{t+1}/N_t)$), calculated using slope method.
alpha:	Median annual population growth rate = e^μ
Last population size:	This is the last running sum value, which acts as the initial population size for the extinction risk estimation.
Approx. conversion factor from running sum to spawners:	If population size where constant, multiplication of the running sum by this factor would yield the number of spawners.

The next section of the file simply repeats the original input data so the input and output information remain coupled. After the original data, the file contains the average age structure. The average age structure is the calculated by pooling all fish in all years and obtaining the fraction of fish of each age class. The next section displays the running sum values versus time, where the running sum is calculated using a user specified method. This table will include corrections to the running sum if “correct for hatchery” or “correct for harvest” are selected (see appendix A). The mean and variance section displays the mean and variance vs. τ . The mean, μ , is estimated as the first order lag (i.e. $\tau=1$), but values at other lags are included for user inspection. The final section of the output file provides the extinction risk estimates. Each row is for a single year in the future, with the first column listing the year, the second listing the probability that the population will reach the QET by that time period, and the next two columns describing the upper and lower @ confidence intervals. The confidence level, @, is specified by the user.

Outfilename_inputfilename_NN File

The *Outfilename_inputfilename_NN* file contains population specific data for plotting $\ln(N_{t+1}/N_t)$ Vs time. The value of $\ln(N_{t+1}/N_t)$ contains any optional corrections for hatcheries or spawners as outlined in table 2 and Appendix A.

Viability Criteria Overview

SimSalmon uses a method of setting viability criteria described by McElhany and Payne (in prep). Basically, a population would be considered viable if it exceeds some minimum combination of growth rate (μ) and population size (N_{min}) over a given period of time. The growth rate and population size combinations are set based on user specified acceptable extinction risk criteria and a data driven estimate of the process variance, σ^2 . The viability criterion is actually a curve, with the region above the curve describing a viable population and the region below the curve describing a non-viable population (Figure 3). A constraint is applied to the criteria that the observed value of the growth rate, μ_{obs} , cannot be less than zero.

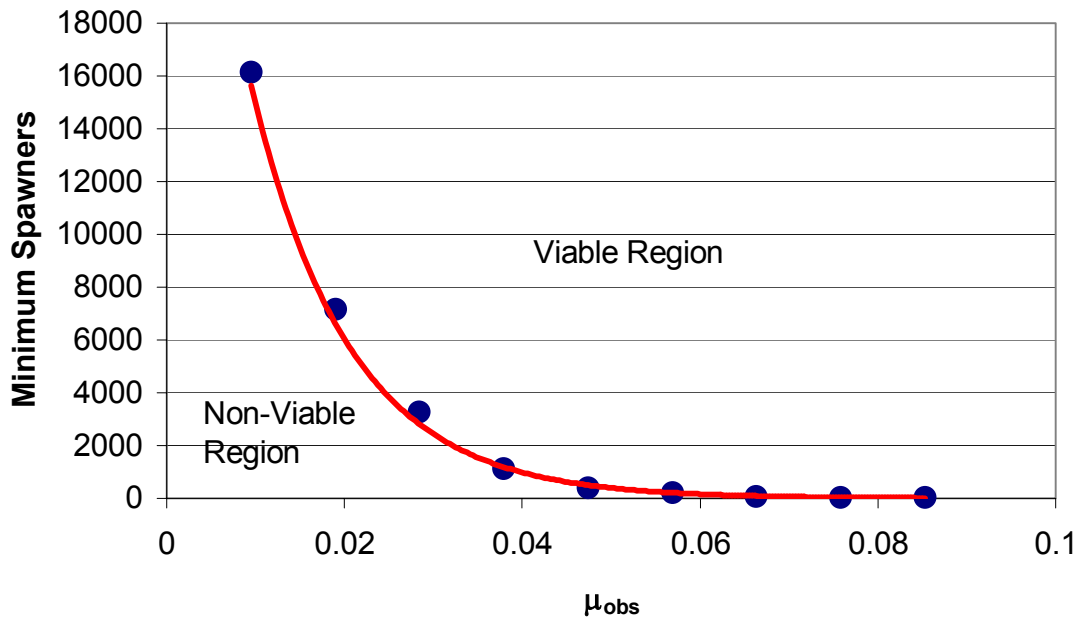


Figure 3: Example of viability criteria. The value of μ_{obs} is the observed point estimate of $\text{mean}(\ln(N_{t+1}/N_t))$. The minimum size is estimated as described in the text.

The approach for setting the minimum size is driven by concern about the impact of environmental variation. All else being equal, a small population has a higher probability of extinction than a large population. The minimum size is estimated as a function of the confidence interval on μ_{obs} , an empirical estimate of the process variance, a user specified acceptable extinction risk and the application of a “hockey stick” type extinction model. The hockey stick population dynamics model projects stochastic exponential growth with a ceiling on total reproductive population size, k . The minimum population size is set as the initial population that just produces an acceptable extinction risk after projecting the hockey stick model into the future. The growth rate parameter of the model is set as the lower confidence interval of μ_{obs} (eq. 8). The user specifies the level of the confidence interval (e.g. 95%). The variance parameter of the model is the empirical estimate based on the historic time series data. The population ceiling, k , is set at the initial population size. Thus, the program estimates the minimum population size under the scenario that the minimum population size is also the population ceiling. This

effectively allows the minimum population size estimate to also be an estimate of minimum carrying capacity or a minimum escapement threshold. We can seldom estimate with confidence the carrying capacity of a population and this approach provides a conservative estimate of the minimum population size, since a population constrained by a low ceiling has a higher extinction risk than a population without a ceiling.

Extinction risk using the model is estimated as the probability that a population starting at some initial population size, N_0 , declines to the quasi-extinction threshold (QET) within a given time horizon. Because the model includes a population ceiling, the extinction risk cannot be estimated using equation 2. The extinction risk is instead estimated by simulating the population projection process many times to produce many population trajectories, and then calculating the fraction of simulated population trajectories the contain populations sizes below QET within the specified time period. To simulate a single trajectory, the population starts at N_0 , and then the population is projected forward by iteratively applying the equation

$$N_{t+1} = N_t e^r \quad \text{if } N_t \leq k$$

$$N_{t+1} = k e^r \quad \text{if } N_t > k$$

$$\text{where } r \approx \text{Normal}(\mu, \sigma).$$

The minimum size is found by a search algorithm that looks for an initial size that generates near exactly the acceptable risk. As noted above, the initial size is also assumed to be a ceiling on the number of spawners. The algorithm starts by estimating the extinction risk for a population with initial size at both lower and upper bounds. The lower bound. If a population with an initial size at the lower bound has an extinction risk below the acceptable risk, the minimum size is reported as less than the lower bound. If a population with an initial size at the upper bound has an extinction risk above the acceptable risk, the minimum size is reported as greater than the upper bound. If the initial size yielding the acceptable risk is between the upper and lower bound, the program uses a search algorithm that repeatedly tests the extinction risk for an initial population size between the last two best estimates. This repeated mid-point search goes through 12 iterations before stopping on a value reasonably close to the initial size with the acceptable risk.

The process variance, σ^2 , is estimated using the Holmes approach and applies to the ratios of the running sum, N . The population growth rate and minimum population sizes are all calculated based on running sums. The minimum size value is estimated as the initial running sum population size that yields an acceptable risk. However, when assessing the viability of a particular population, SimSalmon evaluates the minimum size as the geometric mean of the running sum values over the time period evaluated for viability. Since most the values of μ for the viability criteria are near zero, median population size should remain approximately constant for a viable population. As projected population size is expected to show a lognormal distribution, the geometric mean population size should be approximately equal to the initial population size. Evaluating the geometric mean as the minimum size has the advantage of using multiple years running sum data. The alternative is to simply look at the first running sum of the time series to be evaluated for viability and determine if it exceeds the minimum size criteria. The geometric mean is only an approximation, but given the uncertainty associated with the minimum size estimate itself, this is likely of little consequence.

Viable Criteria Option of Dennis-Holmes Parameters and Extinction Risk Sub-module

Documentation under construction...

Viable Criteria Test Sub-module

Documentation under construction...

The screenshot shows the 'Viable Criteria Test' window within the SimSalmon application. The window has a menu bar with 'File', 'Edit', 'Analysis', and 'Help'. The title bar reads 'SimSalmon' and 'Viable Criteria Test'. The main area contains several input fields and checkboxes. At the top, there are two checkboxes: 'Input dataset includes harvest column' and 'Correct for harvest'. Below these is a checkbox for 'Correct for hatchery fish' followed by a text field for 'Effectiveness of Hatchery Spawners' with the value '0.5'. The main text reads: 'The acceptable risk for the viability criteria is a 5 percent chance of declining to a running sum of 1 fish in 100 years. Number of Replicates for simulation:'. Below this is a 'Confidence Level (%)' field. There are two 'Browse...' buttons for 'Data Input File' and 'Data Output File'. A 'Run!' button is centered below these. At the bottom, there is a 'Comments' label and a large text area for input.

Figure 4: Viable Criteria Test sub-module input screen.

Extinction Risk Table Sub-Module

The Extinction Risk Table sub-module uses the Dennis method to estimate extinction risk. The sub-module outputs a text file that is a table of extinction risk as a function of μ and σ^2 (Table 8). The input screen is shown in Figure 5 and the input variable descriptions are in Table 9.

Table 8: Example output file from Extinction Risk Table sub-module.
The probability of going from an initial size of 500.0 to 20 in 100 years.
The extinction model is the Dennis et al diffusion model of stochastic exponential growth.

Sigma2\Mu	-0.1	-0.05	0	0.05	0.1
0.01	1	0.973142	0.001287	1.02E-14	1.10E-28
0.02	0.999999	0.926275	0.022841	9.49E-08	1.05E-14
0.03	0.999979	0.895694	0.06311	1.96E-05	4.79E-10
0.04	0.99984	0.87539	0.107521	2.80E-04	1.02E-07
0.05	0.999449	0.861297	0.150002	0.001378	2.56E-06

SimSalmon

File Edit Analysis Help

Extinction Risk Table

Initial population size:

Quasi-extinction threshold:

Time Horizon:

Estimate extinction risk for mu from to by .

Estimate extinction risk for sigma2 from to by .

Comments

Figure 5: Extinction Risk Table input screen.

Table 9: Description of input parameters for Extinction Risk Table sub-module.

Input Field	Description
Initial population size	Initial size of the population. This is not an age structured model, as it only looks at the relationship between N_t and N_{t+1} . For salmon, a running sum input may be appropriate.
Quasi-extinction threshold	The size at which the population is functionally extinct. Again, this is not an age structured model, as it only looks at the relationship between N_t and N_{t+1} . For salmon, a running sum input may be appropriate.
Time horizon	The time horizon for evaluating extinction risk.
Estimate extinction risk for mu	The program runs through a 'for loop' and estimates extinction risk for multiple values of μ . The user specifies the start, stop and increment values of the for loop.
Estimate extinction risk for sigma2	The program runs through a 'for loop' and estimates extinction risk for multiple values of σ^2 . The user specifies the start, stop and increment values of the for loop.
Browse	The browse button provides a standard file selection dialog box to specify the output file name and location.

Minimum Size Table Sub-module

The minimum size table sub-module estimates the minimum viable population size using the algorithms described in the viable criteria module. The minimum size is the initial population size that just yields an acceptable extinction risk. Extinction risk is estimated using the exponential growth with a ceiling simulation model (i.e. 'hockey stick model'), where the ceiling is set as the initial population size. The extinction risk for a given parameter combination is estimated by simulating a number of trajectories of the appropriate (i.e. time horizon) length. The stochastic process is simulated by a random draw of N_{t+1}/N_t for each year of the simulated trajectory. The extinction risk is calculated as the fraction of trajectories that go below the QET.

The minimum size is found by a search algorithm that looks for an initial size that generates near exactly the acceptable risk. As noted above, the initial size is also assumed to be a ceiling on the number of spawners. The algorithm starts by estimating the extinction risk for a population with initial size at both lower and upper bounds. The lower bound is QET plus 10 and the upper bound is 100,000. If a population with an initial size at the lower bound has an extinction risk below the acceptable risk, the minimum size is reported as the lower bound. If a population with an initial size at the upper bound has an extinction risk above the acceptable risk, the minimum size is reported as the upper bound. If the initial size yielding the acceptable risk is between the upper and lower bound, the program uses a search algorithm that repeatedly tests the extinction risk for an initial population size between the last two best estimates. This repeated mid-point search goes through 12 iterations before stopping on a value reasonably close to the initial size with the acceptable risk.

This is not an age structured model, as it only examines the relationship between N_t and N_{t+1} . Therefore, for salmon, input values that refer to running sum values may be appropriate. Likewise, the output may be interpreted as some sort of running sum. Sample output is shown in Table 10. The input screen is shown in Figure 6, with a description of the input fields in Table 11.

Table 10: Example output from Minimum Size Table output.

The minimum size is the initial running sum that just produces an acceptable extinction risk.

The extinction model is stochastic exponential growth with a ceiling equal to the initial population size (i.e. minimum size).

Acceptable risk is a 5.0 percent probability of declining to a running sum of 20 fish in 100 years.

Sigma2\Mu	-0.05	0	0.05	0.1
0.01	16826	183	36	36
0.02	37574	379	61	36
0.03	59691	819	85	36
0.04	99085	1577	134	61
0.05	100000	2285	183	85

SimSalmon

File Edit Analysis Help

Minimum Size Table

Acceptable risk is percent probability of declining to fish in years.

Estimate minimum size for mu from to by .

Estimate minimum size for sigma2 from to by .

Number of replications for simulation:

Minimum size analysis range: to

Browse...

Run!

Comments

Figure 6: Minimum Size Table sub-module input screen.

Table 11: Description of input parameters for Minimum Size Table sub-module.

Input Field	Description
Acceptable Risk	The initial population with exactly this level of risk is considered the minimum viable population. The percent probability is a value between 0 and 100, with typical values of 90, 95 or 99. The next field is the quasi-extinction threshold (QET) below which the population is considered functionally extinct. Since this is not an age structured model, this will often be some sort of running sum parameter. The number of years describes the time horizon for evaluating the extinction risk.
Estimate minimum size for mu	The program runs through a 'for loop' and estimates minimum size for multiple values of μ . The user specifies the start, stop and increment values of the for loop.
Estimate minimum size for sigma2	The program runs through a 'for loop' and estimates minimum size for multiple values of σ^2 . The user specifies the start, stop and increment values of the for loop.

Number of replicates for simulation	This input parameter describes the number of simulated trajectories per parameter combination. A value of 2,000 trajectory simulations is generally sufficient, but fewer trajectories will run make the program run faster.
Minimum size analysis range	This is the range over which SimSalmon will search for the minimum size. The smaller the range around the true value, the more precise will be the estimate. If the minimum size is above or below the range, the program output the user specified upper or lower bound as the minimum Size.

Simulate Trajectories Module

Simulate Trajectories Overview

This module simulates population trajectories over time for a partially age-structured population, given a starting population size and a set of parameters that define how the population grows. The module allows the user to simulate population trajectories using 3 models of density-dependent growth (Ricker, Beverton-Holt, and a "hockey-stick" model) or a model of density-independent growth (the Dennis model). In addition it is possible to simulate cyclical regime shifts, declining or increasing carrying capacity, and to add normally-distributed "observation error" as well as lognormal or autocorrelated "process error".

We make the distinction between "process error" and "observation error". *Process error* is random variation in population growth that may be due to environment, demography, genetics or any of a host of other causes not directly explained by the mechanistic model of population growth being used. *Observation error* is measurement error, i.e. in this case it represents our inability to count the spawners that are actually present. An important difference between the two is that process error in a given year will influence later years, i.e. it may accumulate over time, whereas observation error will not. Observation error has no effect on the biological process--it simply makes it harder for us to divine what is happening.

In each year simulated, the program goes through the following steps:

1. Calculate the current year's total spawners, S_{yr} , from recruitment in previous years.
2. *Optional*: If S_{yr} is greater than a specified spawner ceiling, reduce the total spawners to that ceiling (in the case where the user wants to limit the maximum number of effective spawners).
3. If S_{yr} falls *below* the depensation threshold, set recruitment to 0 this year (note that the population may not go extinct if other cohorts are still alive that fall above the threshold).
4. If S_{yr} falls *above* the depensation threshold, calculate total recruits for this year from S_{yr} based on the spawner-recruit function chosen by the user (see below).
5. Add process error, if it was not already introduced in step 4 (see below).

In step 1, the total spawners in the current year is calculated as the sum of the spawners at each age in the current year. The spawners at any given age a are a proportion A_a of the total recruits a years ago:

$$S_{yr} = \sum_{a=first.age}^{max.age} S_{yr,a} = \sum_{a=first.age}^{max.age} R_{yr-a} A_a ,$$

where S_{yr} is the total number of spawners in year yr , and R_{yr-a} is the recruits in year $yr-a$. A_a is the (fixed) proportion of spawners that spawn at age a .

In step 4, total recruits R_{yr} are calculated as $R_{yr} = f(S_{yr})$, where the function $f(S_{yr})$ is specified by the user.

Note that this is not a fully age-structured model, since only spawners are tracked. Also, the age structure A of spawning fish is fixed over time. Further, in the density-dependent section of the model, we are not specifying where in the life cycle the density-dependence takes place.

DENSITY-INDEPENDENT POPULATION GROWTH: THE DENNIS MODEL

In this model, “process error” is introduced by the e^x term. A new x is drawn for each year from a normal distribution with the specified mean and variance.

$$R_{yr} = e^x S_{yr}, \quad \text{where } x \sim N(\mu, \sigma^2).$$

DENSITY-DEPENDENT GROWTH

Three different models of density dependent growth are included in the *Simulate Trajectories* module. They are the Ricker function, the Beverton-Holt function, and a “hockey-stick” function, described below. They are all used in a similar fashion to the Dennis model, except that we replace the random variable e^x with a deterministic function $f(S)$ that specifies a density-dependent relationship between spawners and recruits, and then (optionally) add log-normally distributed process error to each year’s recruitment.

DENSITY-DEPENDENT MODELS

1. Ricker. The Ricker model is one of the most widely-used models of density dependent growth. It has the property that recruitment reaches a maximum at some number of spawners, and may actually decline beyond that maximum as the number of spawners increases. The parameterization used is:

$$R = aSe^{-bS}$$

where R is recruits, S is spawners, a is the initial slope (i.e. the recruits-per-spawner in ideal conditions) and $\frac{a}{b}e^{-1}$ is the maximum possible number of recruits (in our module, the user sets the recruit maximum and b is back-calculated from it).

2. Beverton-Holt. The Beverton-Holt model, another curvilinear model of density-dependent growth, is less flexible than the Ricker model. Recruitment reaches an asymptote at some number of spawners, but does not fall at higher spawner numbers.

$$R = \frac{aS}{1 + \frac{a}{b}S}$$

With this parameterization, a is the initial slope and b is the maximum recruits.

3. Hockey Stick. This is the simplest of the density-dependent growth models. In it, recruitment rises linearly with spawner abundance up to some maximum number of effective spawners S_{max} , above which R/S is constant.

$$R = aS, \text{ if } S \leq S_{max}$$

$$R = aS_{max}, \text{ if } S > S_{max}$$

PARAMETERS THAT CONTROL "ERROR" ADDED TO THE SIMULATIONS

Process Error

The Dennis model implicitly includes process error, because the recruit:spawner ratio is modified each year by a log-normally distributed random variable whose mean and variance are given by the user.

The Ricker, Beverton-Holt and hockey-stick models are deterministic models that do not include process error. They will yield the same result, run after run, if the process error mean and variance are both set at 0. Recruits are calculated each year as if there was no error, and the result is then multiplied by a term that introduces a lognormal error. If $R_{t,no,error}$ is the recruits for a given year *without* process error, then the year's recruits *including* process error are:

$$R_{yr,error} = R_{yr,no,error} \cdot e^x, \text{ where } x \sim N(\mu, \sigma^2), \text{ and } \mu \text{ and } \sigma^2 \text{ are specified by the user.}$$

Autocorrelated Process Error

"Autocorrelated error" means that the population growth error in a given year is more likely to be like the errors in the years preceding it than would be expected if it was drawn from a random normal sample. This tends to create runs of bad or good years, which can increase the magnitude and length of swings in population size.

The program calculates a variance-covariance matrix based on the process error variance and the autocorrelations specified by the user. This is a hair-raising process involving someone named Cholesky and a matrix that must be positive definite. The program will

create a vector of autocorrelated errors from a normal distribution with the variance-covariance matrix it calculated, and apply them as described in the Process Error section. The process is described in more detail above (See Modified Dennis: Spawner/Spawner).

Observation Error

Observation error only added once the trajectory is complete. It is assumed to be normally distributed with a mean of 1, and it is added multiplicatively (like process error, above):

$$R_{yr,ObsError} = R_{yr,no.ObsError} \cdot s, \text{ where } s \sim N(1, \sigma^2), \text{ and } \sigma \text{ is specified by the user.}$$

Regime shifts

The simulator allows the population to alternate between two different sets of growth parameters throughout the simulation, as might happen if the environment was cycling between favorable and unfavorable conditions. We allow for two sets of parameters, one for each regime. The length of the regimes and the year in which the second regime starts may be set by the user.

Changing carrying capacity

We simulate a changing carrying capacity with the density-dependent models by changing the maximum number of recruits each year. The recruitment ceiling may increase or decrease linearly throughout each regime. Each year's ceiling, $Rmax_{yr}$, is calculated before the recruitment is calculated, as:

$$Rmax_{yr} = m \cdot (yr - startyr) \cdot Rmax_{init} + Rmax_{init}$$

where $Rmax_{init}$ is the initial recruit ceiling as specified by the user, m is the slope of the change in the ceiling, and yr and $startyr$ are the current and initial years, respectively.

How to use the *Simulate Trajectories* module

The output from this module is a series of text files, one for each population trajectory. These files in the format input into the other SimSalmon modules. The files are organized in the following columns:

Year	Year of simulation
Abund	Spawner abundance for that year
fracWild	fraction of the year's spawners that are wild-origin (not hatchery) fish
age1	These columns give the fraction of the spawners by age
age2	...

etc. (to max age)...

SimSalmon

File Edit Analysis Help

Simulate Trajectories

Maximum Age at Reproduction: 5

Number of Trajectories: 10

Age Distribution of Spawners

Age1	Age2	Age3	Age4	Age5
0.0	0.01	0.19	0.57	0.23

Initial Population Size

Year1	Year2	Year3	Year4	Year5
1000	1000	1000	1000	1000

Depensation Threshold: 10

Start Time: 1

Stop Time: 150

Regime 1 Length: 150

Regime 2 Length: 0

☐ Regime 2 Start (check for random start) 0

Calculation Type: Hockey Stick

R1 Initial Slope R/S: 1 R2 Initial Slope R/S: 1

R1 Maximum Spawners: 1,000 R2 Maximum Spawners: 1,000

R1 slope of max spawners ceiling: 0 R2 slope of max spawners ceiling: 0

R1 Mean of log-normal process error: 0 R2 Mean of log-normal process error: 0

R1 Variance of log-normal process error: 0 R2 Variance of log-normal process error: 0

☐ AutoCorrelation (Max. Lag) 3

Lag1: 0.0 Lag2: 0.0 Lag3: 0.0

☐ Observation Error Std. Dev. 0 from years 6 to year 50

Base Output File: Browse... Run!

Figure 7: Input screen for Simulate Trajectories.

Top Pane Parameters

Maximum age at reproduction: This determines the number of boxes for initial population size and spawner age distribution.

Age Distribution of Spawners: This vector of numbers fixes the proportion of a cohort that will spawn at a particular age. The values do not change during the simulation. For example, if you set the age 3 box to 0.27, then 27% of every cohort will spawn at age 3. The boxes must sum to 1. Note that the proportion of spawners observed on the spawning grounds of a given age (what you see in the output file) will vary over time, as larger and smaller cohorts pass through the population.

Initial Population Size: This is a vector, rather than a single number, of the total spawners in years 1 to year a_{max} , where a_{max} is the maximum age of reproduction. A vector is needed because spawners in year x are calculated from spawners in years $(x - \text{age of first reproduction})$ to $(x - \text{age of last reproduction})$.

Number of Trajectories: The number of trajectories, and thus also the number of files that will be created when you push Run.

Depensation Threshold: If, in a given year, there are fewer spawners than the depensation threshold, reproduction is set to zero for that year. If the threshold is set to 1, depensation will not affect population growth.

Start Time and Stop Time: These determine the length of each trajectory in years. Start time should be set to 1.

BOTTOM PANE PARAMETERS

Regime 1 Length: Sets the length, in years, of regime 1.

Regime 2 Length: If set to zero, then all regime 2 parameters are ignored and the simulations run only under regime 1 parameters. Otherwise, this sets the length of regime 2, in years.

Regime 2 Start: The simulation always starts in regime 1. It switches to regime 2 and starts cycling (unless the regime 2 length is set to 0), beginning in the year specified here. If the “check for random start” box is checked, the start year for regime 2 will be randomly chosen (within the course of regime 1) for each trajectory.

Calculation Type:

This is a drop-down menu that allows the user to choose between different spawner-recruit models. *Each model has a set of associated parameters arranged in two columns below the drop-down menu. On the left are regime 1 parameters, which have names beginning with R1. The right-hand column parameters, whose names begin with R2, are the regime 2 equivalents, and may be ignored if regime 2 length is set to zero.*

The models are:

1. Spawner:Spawner. This selects Dennis model calculations. This is the model used in the rest of SimSalmon. Its assumptions are discussed above. The user sets the mean (**Mean Spawner:Spawner**) and variance (**Variance Spawner:Spawner**) of the population growth rate $N_t/N_{t+\tau}$. If the **Max Spawners (ceiling)** box is not checked, there is no limit on the possible number of spawners. If the box is checked, the number of fish that spawn in each year will be limited to the number entered.
2. Ln(Spawner:Spawner). This also selects Dennis model calculations, the only difference being that log units are used here for the mean and variance, i.e. the user should set the mean and variance of $\ln(N_t/N_{t+\tau})$.
3. Ricker. The Ricker parameters are the **Initial Slope of R/S** (a , above), and the **Maximum Recruits** (from which the Ricker b parameter, above, is calculated).

The **Slope of max recruits ceiling** parameter may be set positive or negative so that the maximum recruits number increases or declines linearly from the initial ceiling set by Maximum Recruits. If the slope is set to -0.05 , for instance, the ceiling will decline at 5% per year. The default is zero (no change).

Process error may be added by specifying the **Mean of log normal process error** and the **Variance of log normal process error**. Note: to choose the error so that the *median* error will be zero, the user should set the process error mean to 0, since $e^0 = 1$ (however, the *mean* error will be greater than zero, since the mean of a log-normal distribution is $e^{\mu^2/2}$).

4. Beverton-Holt. In this case the **Initial Slope of R/S** is the a parameter of the model parameterization used (see above), and the **Maximum Recruits** is the b parameter. Other parameters are the same as for the Ricker function.
5. Hockey stick. **Initial Slope of R/S** is a (above). **Maximum Spawners** is S_{max} (above). Other parameters are the same as for the Ricker function.

Autocorrelated process error

If the **Autocorrelation (max. lag)** box is checked, the process error will be autocorrelated according to the values given. The maximum lag determines the length of the span of past years that influence the current year, and the program creates a set of boxes labeled **Lag1, Lag2, Lag3...** up to the maximum lag specified. These boxes must contain the correlations between pairs of years separated by 1, 2, 3...etc. years, and the process error variance must be set to some positive number. It is possible to specify a correlation structure that doesn't make sense, in which case the Cholesky matrix decomposition process will fail and the program will complain.

Observation Error

If the **Observation Error Std. Dev.** box is checked, normally distributed observation error will be added to the finished trajectory, as explained above. The error distribution will have a mean of 1 and a standard deviation specified in the box to the right of the label. Error is only added to those years specified (so that the true trajectory can be known when evaluating delisting decisions).

BASE OUTPUT FILE

The output files you create will be placed in the directory you specify here. For each population trajectory simulated, a number is added to the end of the file name you specify, e.g. if you name your base file "myoutput", the program will create files named *myoutput_1*, *myoutput_2*, *myoutput_3*, etc.

The program will also create and save one text file called *myoutput_par* in the same directory, that contains a list of the common parameter values used to create the trajectories. The *myoutput_par* file is used by the *Evaluate Criteria* module to keep track of the parameter values associated with groups of trajectories.

Evaluate Criteria Module

Evaluate Criteria Overview

This module exists for the purpose of evaluating the performance of the de-listing criteria. The idea is to produce a large set of population trajectories with known characteristics, to make a decision whether or not to de-list each population based on the early years of its trajectory, and then to use the end of each trajectory to evaluate the fate of the population after the de-listing decision was made. To implement this strategy, we divide each trajectory into 3 sections. Sections 1 and 2 may overlap.

1. A *Variance Estimation Period*, in which the variance of the growth rate is estimated.
2. A *Decision Period*, in which the mean growth rate and geometric mean population size are estimated and a decision is made to de-list or not, based on the criteria given.
3. An *Evaluation Period*, in which the subsequent performance of the population is evaluated.

The results are compiled into a table that shows how the criteria performed. This allows us to answer questions such as:

- What proportion of the trajectories that were de-listed subsequently went extinct during the evaluation period?
- How fast would a population have to grow (given some variance of the growth rate) to be de-listed?
- How do trade-offs between the mean and variance of the growth rate affect the probability of de-listing populations?
- How do such potential complications as density-dependent growth, regime shifts in ocean productivity, autocorrelated process errors, and observation error affect the probability of de-listing?

Essentially, this module wires together several other modules. The variance of the growth rate is estimated using the “Dennis Holmes parameters and extinction risk” module (on the Modified Dennis menu). The de-listing decision, based on growth rate from the decision period and a variance estimate derived from the variance estimation period is made by the “Viable Criteria Test” module (also on the Modified Dennis menu).

How to use the *Evaluate Criteria* module

SimSalmon

File Edit Analysis Help

Evaluate Criteria

Add file/s to batch analysis... Remove selected file

Output File Path . Browse...

The acceptable risk for the viability criteria is a 5 percent chance of declining to
a running sum of 10 fish in 100 years. Number of Replicates for simulation: 1,000

Variance estimation period 11 50

Decision period 31 50

Evaluation period 51 150

Confidence level (%): 95

Comments

Run!

Figure 8: Input screen for Evaluate Criteria module.

Before you begin, you should have created a set of trajectories using the *Simulate Trajectories* module. It is easiest if this set of trajectories is in its own directory.

Add file/s to batch analysis

The first step is to enter the files in the **Add file/s to batch analysis** box. Note that the *Simulate Trajectories* module creates a text file with the same base name as the other files, but ending with the suffix “_par”. This file contains a list of parameter values that created the trajectories. It should remain in the directory with the trajectories but should *not* be included in the box.

Output File Path

The **Output File Path** will be the directory where the single output file is put. The output file is a small tab-delimited text file suitable for input into MS Excel. Its name will begin with the base file name of the last input file, and end in *_CEVAL*.

acceptable risk

The **acceptable risk** statement sets the risk for the viability criteria. A population will be de-listed if it meets the criteria given.

Number of replicates for simulation

Number of replicates for simulation controls how many iterations are done in the Viable Criteria Test module. The criteria consist of two parts: a requirement that the growth rate be positive, and a requirement that the minimum geometric mean size of the population be greater than some minimum size. The simulations are done to choose the minimum size.

Variance Estimation Period, Decision Period and Evaluation Period

The boxes for **Variance Estimation Period**, **Decision Period** and **Evaluation Period** control the range of years included in each period (*from* the year in the lefthand box *to* the year in the righthand box). The variance estimation period may overlap with the decision period. By default the variance estimation period begins at year 11, as a small buffer to avoid initial transient changes in age distribution.

Confidence Level

The **Confidence Level** is really part of the acceptable risk statement: for example: we want to be 95% confident that there is a 5% chance of declining to a running sum of 10 fish in 100 years.

OUTPUT

A single tab-delimited text file is created as output. It has three sections:

- 1) The top section is simply a copy of the *_par* file that was created when you made the trajectories. It lists the parameters you used to create the trajectories.
- 2) The middle section contains three summary tables. The first is a 2X2 table labeled SUMMARY TABLE that shows the number of trajectories that went Extinct/Not Extinct against the number of trajectories that were Delisted/Not Delisted.

To its right is a table labeled END, which stands for “**Extinct-Not Delisted**”. It shows the subset of runs that went extinct and were not delisted. It shows why they failed to be delisted. Either the initial size was lower than the minimum size specified by the criteria ($Init < Min$), or the growth rate was lower than zero ($\mu < 0$), or both.

The table labeled NEND does the same thing for the subset of runs that were “**Not Extinct-Not Delisted**.”

- 3) The bottom of the file contains information for each trajectory. A description of the columns follows:

Table 12: Output fields for Evaluate Criteria module.

Column Heading	Contents of column
Traject #	Corresponds to the suffix of the trajectory’s filename
Mu VEP	Mean population growth rate, μ , during the variance estimation period.
Sigma2VEP	Variance, σ^2 , of $\ln(N_{t+1}/N_t)$ during the variance estimation period.

mu DP	Mean population growth rate, μ , during the decision period.
MuLowerDP	Lower 95% confidence limit on mu DP.
ArMeanDP	Arithmetic mean of the running sum during the decision period.
GeoMeanDP	Geometric mean of the running sum during the decision period.
HarmMeanDP	Harmonic mean of the running sum during the decision period.
Mu EP	Population growth rate, μ , during the evaluation period.
ArMeanEP	Arithmetic mean of the running sum during the evaluation period.
GeoMeanEP	Geometric mean of the running sum during the evaluation period.
HarmMeanEP	Harmonic mean of the running sum during the evaluation period.
MinSize	Minimum running sum above which the geoMeanDP must fall for a viable population according to the criteria.
MinSizeLow	MinSize calculated using the lower confidence limit on the variance estimate.
MinSizeUpper	MinSize calculated using the upper confidence limit on the variance estimate.
Delisted?	Was the population delisted?
Extinct?	Did the running sum dip below the quasi-extinction threshold at any time during the decision or evaluation periods?

Recruits per Spawner Module

Recruits per Spawner Overview

The recruits per spawner analysis is based on the simple equation

$$N_t = \sum_{i=1}^{\text{max. age}} N_{t-i} R_{t-i} A_i, \quad \text{eq. 1}$$

where N_t is the number of spawners at time t , R_t is the spawners per spawner at time t , A_i is the average proportion of spawners that are age i and maxAge is the maximum age at maturation. The spawners per spawner parameter, R_t , is a random variable that introduces environmental stochasticity into the model. In the basic model, R is assumed to be lognormally distributed and the two shape parameters of the distribution can be estimated from a time series of abundance. More complex scenarios involving temporal autocorrelation or trend in R can also be explored using the model.

The basic simulation protocol involves choosing an initial number of spawners for years one through maxAge. The number of spawners is then projected forward in time by looping through the equation above. This process is repeated 100's or 1,000's of times and the proportion of population trajectories that go extinct by a given period of time is reported as the probability of extinction for that time horizon. An example trajectory calculation is shown in Table 13.

Table 13: This table shows how spawner abundance is calculated for each time period. The initial spawner abundance is specified as 5,000 for the first five years. The recruits per spawner (R) is randomly drawn each year from a lognormal distribution. The distribution parameters of R used in this example were mean = 1, variance 0.5. The age structure in the model is constant such that zero recruits mature at age 1, 5% at age 2, 39% at age 3, 54% at age 4, and 2% at age 5. The number of spawners in a given year is found by adding up the fish that mature in that year, as indicated by the cells of the table highlighted in yellow.

Time	N = Spawners	R = Recruits Spawner	R * N = Recruits	Age Distribution				
				Age 1	Age 2	Age 3	Age 4	Age 5
				0.00	0.05	0.39	0.54	0.02
Init. 1	5000	0.469	2347	0	117	915	1267	47
Init. 2	5000	0.311	1554	0	78	606	839	31
Init. 3	5000	0.895	4475	0	224	1745	2416	89
Init. 4	5000	2.671	13355	0	668	5208	7211	267
Init. 5	5000	1.560	7802	0	390	3043	4213	156
1	3299	0.131	433	0	22	169	234	9
2	8046	0.307	2467	0	123	962	1332	49
3	10365	1.886	19545	0	977	7623	10554	391
4	4773	1.187	5666	0	283	2210	3060	113
5	2329	0.409	952	0	48	371	514	19
6	9247	0.299	2761	0	138	1077	1491	55

The model utilizes a user specified “quasi-extinction” threshold. If the average number of spawners/year drops below the threshold for maxAge years in a row, then the population is considered extinct. (That is if $\sum_{i=t}^{t+\max Age} n_i \leq \text{threshold} * \max Age$, the population is considered extinct at time $t + \max Age$.)

The model utilizes a user specified depensation threshold (henceforth, “threshold”). Thus, the depensation threshold is not equivalent to the “quasi-extinction” threshold of Dennis type models.

Density dependence is currently incorporated into the model by including an optional depensation threshold and/or ceiling on the total number of spawners. If the number of spawners drops below the depensation threshold, the recruitment, R , is zero for that year. Although recruitment for a given year may be zero, it does not mean that the population is necessarily extinct. There may still be individuals out in the ocean that will return to spawn the next year. A population is only considered extinct if it drops below the quasi-extinction threshold. If a ceiling is included in the simulation, when $N > K$, then the program sets $N = K$, where K is the carrying capacity or ceiling.

Confidence intervals can be determined for the extinction risk estimate by bootstrapping using a method analogous to that used in Dennis et al. It is easiest to work with $\ln(\text{spawners/spawner})$ rather than R . Since R is lognormal, the distribution of $\ln(\text{spawners/spawner})$ is normal. The mean, μ , and variance, σ^2 for $\ln(\text{spawners/spawner})$ are related to the mean, ω , and variance, ψ^2 , of the lognormally distributed R as follows,

$$\omega = e^{\left(\mu + \frac{\sigma^2}{2}\right)}, \quad \text{eq. 2}$$

$$\psi^2 = (e^\mu)^2 * e^{\sigma^2} * (e^{\sigma^2} - 1) \quad \text{eq. 3}$$

The mean of the $\ln(\text{spawners/spawner})$, μ , is distributed

$$N\left(\mu, \frac{\sigma}{\sqrt{\text{sampleSize}}}\right). \quad \text{Eq. 4}$$

The variance of $\ln(\text{spawners/spawner})$, σ^2 , is distributed

$$\text{Chi}(\text{sampleSize} - 1) * \frac{\sigma^2}{(\text{sampleSize} - 1)}. \quad \text{Eq. 5}$$

By taking repeated, independent draws from these distributions, then recalculating the extinction risk [after converting from $\ln(\text{spawner/spawner})$ back to R], it is possible to obtain a distribution for the extinction risk. From the distribution of extinction risk, any desired confidence interval can be determined.

Assumptions of the Model

The Model suffers from many of the same assumptions that afflict the Dennis et. al model.

- (1) Stationary Process –If parameters are to be estimated from data, the data must represent observations of a stationary process. Past observations must predict future behavior of the system.
- (2) Spawners/Spawner Lognormal – This is an important assumption to examine when estimating parameters from data. Most commercial statistics packages will test the fit of distributions, look for outlier data points and help test some of the stationarity assumptions. SimSalmon includes some diagnostics to test this assumption.
- (3) All variability and density dependence can be represented in the spawner/early juvenile life-stage.
- (4) Temporal Autocorrelation either does not exist or can be expressed in the autocorrelation analysis.
- (5) Single Population – The program models a single population. The “Calculate Spawners/Spawner” analysis will do some corrections involving data that included immigrant fish (e.g. hatchery fish). However, the projection used in the extinction risk calculation does not include immigration. This can be an especially troublesome assumption in estimating parameters from a very small population because a very few immigrants can really mess up the spawners/spawner ratio.
- (6) The age structure is constant and is known perfectly.
- (7) Density dependence can only be represented as a population ceiling and a depensitory threshold.
- (8) No Observation Error – In estimating parameters from data, all observations represent true abundance.

Calculate Spawners/Spawner Sub-Module

The equation for calculating r/s is

$$r / s_t = \frac{\sum_{i=1}^{\max \text{ Age}} N_{t+i} A(i)_{t+i}}{N_t}, \quad \text{eq. 6}$$

where N_t is the number of spawners at time t and $A(i)_t$ is the fraction of age i individuals at time t .

The equation for r/s if first generation hatchery fish are spawning is

$$r / s_{\text{withHatchery}} = \frac{\sum_{i=1}^{\max \text{ Age}} (N_{t+i} - M_{t+1}) A(i)_{t+i}}{N_t - gM_t}, \quad \text{eq. 7}$$

where N_t is the number of spawners at time t , M_t is the number of migrants into the population at time t (e.g. first generation hatchery fish) and $A(i)_t$ is the fraction of age i individuals at time t and g is the relative spawning success of hatchery fish as compared to the progeny of natural spawners. The above equation gives the Natural Return Ratio (NRR) for the population. If $g = 1$, first generation hatchery fish just as effective at

spawning as natural fish. If $g = 0$, first generation hatchery fish can not successful spawn in the wild. For many systems, it is suspected that g is greater than zero but less than one.

INPUT WINDOW

The input window for the Calculate Spawners/Spawner analysis is show in Figure 9. The program can read data either with or without a harvest column as described in the Modified Dennis module. If the data includes harvest, the “calculate pre-harvest spawners” option may be selected. The pre-harvest option calculates pre-harvest recruits per spawner, as opposed to the normal option of spawner recruits per spawner. If the check box “Correct for hatchery fish” is NOT selected, equation 6 is used and the data column containing “proportion_natural_origin_spawners” is ignored. If “Correct for hatchery fish” is selected, equation 7 is used. The “Effectiveness of Hatchery Spawners” parameter sets “ g ” in equation 7 and should be a fraction \geq zero and \leq 1. Use the “Browse...” buttons to set the input and output file names. The user can include comments describing the data in the comments box. These comments will be written as part of the output file.

SimSalmon

File Edit Analysis Help

Calculate Spawner/Spawner

☐ Input data includes Catch column

☐ Calculate pre-harvest spawners

☐ Correct for hatchery fish

Effectiveness of Hatchery Spawners 0.5

Data Input File Browse...

Data Output File Browse...

Run!

Comments

Figure 9: Input screen for Calculate Spawners/Spawner.

OUTPUT

As output, the Calculate Spawners/Spawner analysis gives a table of Spawners, Recruits, Recruits/Spawner) and ln(Recruits/Spawner). If the spawners per spawner can

not be calculated for a given year because of missing data, a value of –99 is placed in the output data line. If the abundance count for a given year is listed as zero, a value of one is substituted to avoid the potentially challenging task of division by zero. As summary information, the output file shows mean r/s, variance r/s, mean ln(r/s) and variance ln(r/s). Confidence intervals are calculated on mean ln(r/s) and variance ln(r/s) by assuming that ln(r/s) is normally distributed and using the standard method of estimating confidence intervals for a normally distributed variable (see any statistics text). The output also lists the average age structure, which is calculated as the total number of fish that spawn at a given age, divided by the total number of spawners in the entire data set. Sample Size is the number of observations of r/s. Lambda in the output file is the average annual rate of increase (or decrease) in the population and is calculated as

$$\lambda = e^{r + \frac{\sigma^2}{2}}, \quad \text{eq. 8}$$

$$\text{where } r = \frac{\text{mean}(\ln(\text{recruits/spawner}))}{\text{mean age at reproduction}},$$

$$\text{and } \sigma^2 = \text{var}(\ln(\text{recruits/spawner})).$$

(Note: “r” in this equation does not equal “recruits”).)

The output also contains the result the several diagnostic analyses to check the assumptions of the basic model. The output is not an exhaustive analysis of whether or not the model is appropriate, but it does provide a preliminary summary of the data. First are tests for linear trend in r/s and linear trend in ln(r/s). The output file list the regression slope and the result of a simple parametric test that the slope of the trend in ln(r/s) is different from zero at alpha = 0.05. The program also tests for temporal autocorrelation in ln(r/s). Autocorrelations are calculated based on the raw ln(r/s) data for $n/4$ lags, where n = span of years with data on r/s. A bootstrap test of significance is provided, by randomly pairing years and observations of ln(r/s), then recalculating the autocorrelation. This is repeated 2,000 times to obtain a distribution of the autocorrelation, which is compared to the observed value of as a test of significance. Probability values that are extremely high (e.g. >0.975) indicate a significant positive autocorrelation and values that are extremely low (e.g. < 0.025) indicate a significant negative autocorrelation. Note that several test are being conducted simultaneously (one for each lag period), and some sort of correction in interpretation of the probabilities should be applied (e.g. Bonferroni correction). The temporal autocorrelation is not independent of any trend in ln(r/s), so the two values should be considered together (e.g. a significant trend will tend to generate a positive autocorrelation at small lags.)

Density independence is also an important assumption of the model. The program outputs the slope of the regression of ln(r/s) vs. spawners. If there is no linear relationship between the number of spawners and ln(r/s) the slope should be zero. However, testing significance of this statistic is challenging. We are currently exploring various bootstrap methods to test for density dependence in a time series.

To look for outliers in the data set, the output displays the residuals of ln(r/s) [i.e. observed ln(r/s) – mean(ln(r/s))] and the jackknifed standardized residuals (a.k.a. external studentized residuals). A standardized residual greater than two indicates a “significant” outlier (see discussion in Dennis et al 1991, or standard statistical texts).

Extinction Calculator Sub-Module

The extinction calculator estimates the probability of extinction using the method described in the Recruits per Spawner Overview section.

INPUT WINDOWS

The calculator can be run either in single populations mode, where values are manually input for every parameter, or in batch mode, where data parameters are input into the extinction calculator directly from a data file. Single population mode is discussed first and Figure 10 shows the input screen.

The screenshot shows the 'SimSalmon' application window with the 'Extinction Probability Batch Analysis' sub-window active. The window has a menu bar with 'File', 'Edit', 'Analysis', and 'Help'. Below the menu bar is a large empty text area. Underneath this are two buttons: 'Add file/s to batch analysis...' and 'Remove selected file'. Below these buttons is a section for file paths. It includes a label 'Output File Path', a 'Browse...' button, and a text field containing 'C:\Program Files\NMFS_NWFSC\SimSalmon\data_out'. Below this is a 'Summary File Name' label and a text field containing 'Untitled'. The main parameter section is titled 'Quasi-Extinction Threshold: Average of 50 spawners/year for "Max(Age at Reproduction)" years'. It contains several input fields and checkboxes: 'Correct for Hatchery Spawners' (checkbox), 'Effectiveness of Hatchery Spawners' (text field with '0.5'), 'Depensation Threshold' (text field with '10'), 'Number of Replicates' (text field with '1,000'), 'Maximum Population Size(Ceiling)' (text field with '5,000'), 'Calculate Confidence Interval (CI)' (checkbox), 'No. of Bootstraps' (text field with '1,000'), and 'Maximum Time Horizon' (text field with '100'). At the bottom of this section is a 'Comments' label and a large text area. A 'Run!' button is located at the bottom center of the window.

Figure 10: Input screen for the extinction calculator.

Maximum Age at Reproduction: This parameter sets the maximum age of spawners. The model assumes that individuals do not survive to spawn multiple years.

Age Distribution of Spawners: This input table will have a number of cells equal to the maximum age at reproduction. The fraction of individuals that spawn at each age is entered into the table. Values must be between zero and one (inclusive) and must sum to one. The model assumes a constant tendency to return at a given age.

Initial Population Size: The initial number of spawners must be specified for a number of years equal to the maximum age at reproduction. Input could be informed by the last years of a population's observed time series or initial population sizes of theoretical interest could be explored. Theoretical explorations will usually involve stipulating the same number of spawners in each initial size year.

Quasi-extinction Threshold: If the population drops below this value, the population is assumed to be functionally extinct. The threshold is described in the General Background section.

Calc Type: Data can be input as either Spawners/Spawner [r/s] or $\ln(\text{Spawners/Spawner})$ [$\ln(r/s)$]. The General Background section describes the functional relationship between r/s and $\ln(r/s)$ when it assumed that r/s is lognormally distributed. Data are often available in the literature based on $\ln(r/s)$ and using this form of input may make comparisons to other models easier. The “Calculate Spawners/Spawner” analysis of SimSalmon reports results based on both r/s and $\ln(r/s)$. The extinction calculator will generally give different results depending on which input is used because observed r/s ratios will not be perfectly lognormally distributed.

Average Spawner:Spawner or Average $\ln(\text{Spawner:Spawner})$: The arithmetic mean of either S/S or $\ln(S/S)$.

Variance Spawner:Spawner or Variance $\ln(\text{Spawner:Spawner})$: The variance of either S/S or $\ln(S/S)$.

Maximum Spawner:Spawner or Maximum $\ln(\text{Spawner:Spawner})$: The user can stipulate a maximum value for either r/s or $\ln(r/s)$. This could be a hypothetical or observed maximum productivity. In the implementation of the model, the distribution of r/s is simply truncated at the user-specified maximum. Because of the truncation, the actual mean and variance of the r/s distribution will be smaller than that specified in the average and variance inputs.

Number of Replicates: The number of simulated trajectories used for calculation of the extinction probability. A thousand replicates are usually sufficient for convergence of the extinction probability.

Depensation Threshold: If the number of spawners drops below this value, no recruits are generated from that brood year.

Maximum Time Horizon: The maximum time in the future at which to calculate the probability of extinction.

Maximum Population Size (Ceiling): The user can specify a maximum number of spawners for the population. If the number of returning spawners in a given year exceeds the ceiling, the number of spawners is truncated to the ceiling value.

Calculate Confidence Interval: The user can choose to calculate confidence intervals based on the method outlined in the General Background section. This requires a computationally intensive bootstrap method, which may be time consuming if you have a slow computer.

No. of Bootstraps: This is the number of bootstrap replicates to use for calculating confidence intervals. A thousand should be plenty.

Sample Size: This is the number of r/s observations in the data set. (Note, it is not the number of years of data). It is used to calculate the confidence intervals on the extinction probability.

Comments: The user can write comments related to the analysis. These comments will be written as part of the output file and are a good place to put information describing the data source and choice of parameters.

Write Trajectories: The user can choose to write the simulated trajectories to the output file. This can use up both CPU memory and storage space, and should only be done if you actually plan to look at them.

Output File: The browse button opens a file chooser dialog box to write the name of the output file.

Run!: This button starts the analysis. While the button remains depressed, the analysis is still in processes. (We are working on a better way to inform the user of the program's status.

EXTINCTION CALCULATOR SINGLE MODE OUTPUT FILE:

The output file is tab-delimited text and can be opened in Excel or any program that can open text files. Note excel has a limit on the number of columns it will display and if it is used to view the output and “Write Trajectories” is selected, only the first 256 trajectories will be shown. If this bothers you, call Bill Gates or use another program that does not have a limit on the number of column.

The first part of the file will list the SimSalmon version number, then the date and time the analysis was conducted. Next the model input parameters are displayed, so the input is linked to the output and it is possible to rerun the analysis if desired. The next section of the output file gives the probability of extinction at each projected year in the future. If “Calculate Confidence Intervals” was selected, the output also contains 90, 95, and 99 percent upper and lower confidence intervals. If “Write Trajectories” was selected, the model then writes out all the simulated spawner trajectories.

BATCH MODE:

The extinction calculator can be run in batch mode to make analysis of multiple input files easier. The input screen is shown in Figure 11. In batch mode, a Calc Spawners per Spawner analysis is first conducted on each input file. The results of the r/s analysis are then used to parameterize an Extinction Probability Calculator analysis. Two

output files are generated for each input file. These output files are given the names FileName.sps and FileName.risk, where “FileName” is the name of the input file. The “.sps” file is the output file from the Calculate Spawners per Spawner analysis and the “.risk” file is the output from the extinction risk calculator.

In addition, two summary files (“.sps” and “.risk”) are generated to the output path with the base name specified in the “Summary Field Name”. These files contain summary information on the files. The “.risk” summary file shows the probability of extinction for each population at each time period. If confidence intervals are specified in the output, the file also shows the 95% C.I. The “.sps” summary file summarizes the spawners/spawner analysis for each population. The fields displayed are show in table 14.

Table 14: Fields in batch mode summary.sps.

Field Name	Description
fileName	name of input file
mean(r/s)	mean recruits/spawner
var(r/s)	variance in recruits/spawner
mean(ln(r/s))	mean ln(recruits/spawner)
var(ln(r/s))	variance in ln(recruits/spawner)
&mean(r/s)	mean r/s calculated from mean ln(r/s) assuming lognormal distribution
&var(r/s)	variance r/s calculated from mean ln(r/s) assuming lognormal distribution
lambda	mean annual rate of increase
sampleSize	number of estimates of r/s
trend(r/s)	slope of linear trend in r/s
trend(ln(r/s))	slope of linear trend in ln(r/s)
isTrendSig	test if slope of ln(r/s) Vs. time is significantly different from zero at alpha = 0.05
isAutocor	test if the lag 1 autocorrelation of ln(r/s) is significantly different from zero at alpha = 0.05
hasOutliers	test if any of the standardized jackknife residuals are > 2
maxAge	maximum age at reproduction
ageDistributi on	mean proportion spawning at each age
recentPopSi ze	population size for last "maxAge" years

A note of caution is in order regarding summary files. In order for the files to easily import into graphing programs, these files do not contain input parameter information. It is important to retain some separate records of the input parameters associated with each summary file.

The Extinction Calculator is passed the mean(ln(r/s)), var(ln(r/s), age distribution, initial population size, and sample size. The initial size is set as the size of the last “maxAge” years of the spawner abundance time series. While batch mode will test for trend or autocorrelation in r/s and show the output in the .sps files, this information is not included in the risk analysis. I patterns in r/s are detected, a separate analysis needs to conducted.

The parameters set in the input window apply to every input file analysis and are identical to the parameters of the same name in the Calculate Spawners per Spawner and Extinction Calculator Single Population Analysis input windows. (Note: “Output File” sets the path for the output files. When using the path chooser accessed by the browse button, you cannot actually choose a folder when you are “inside” the fold, you need to go up a level and simply highlight the folder. This is a quirk of Java, and can cause confusion.)

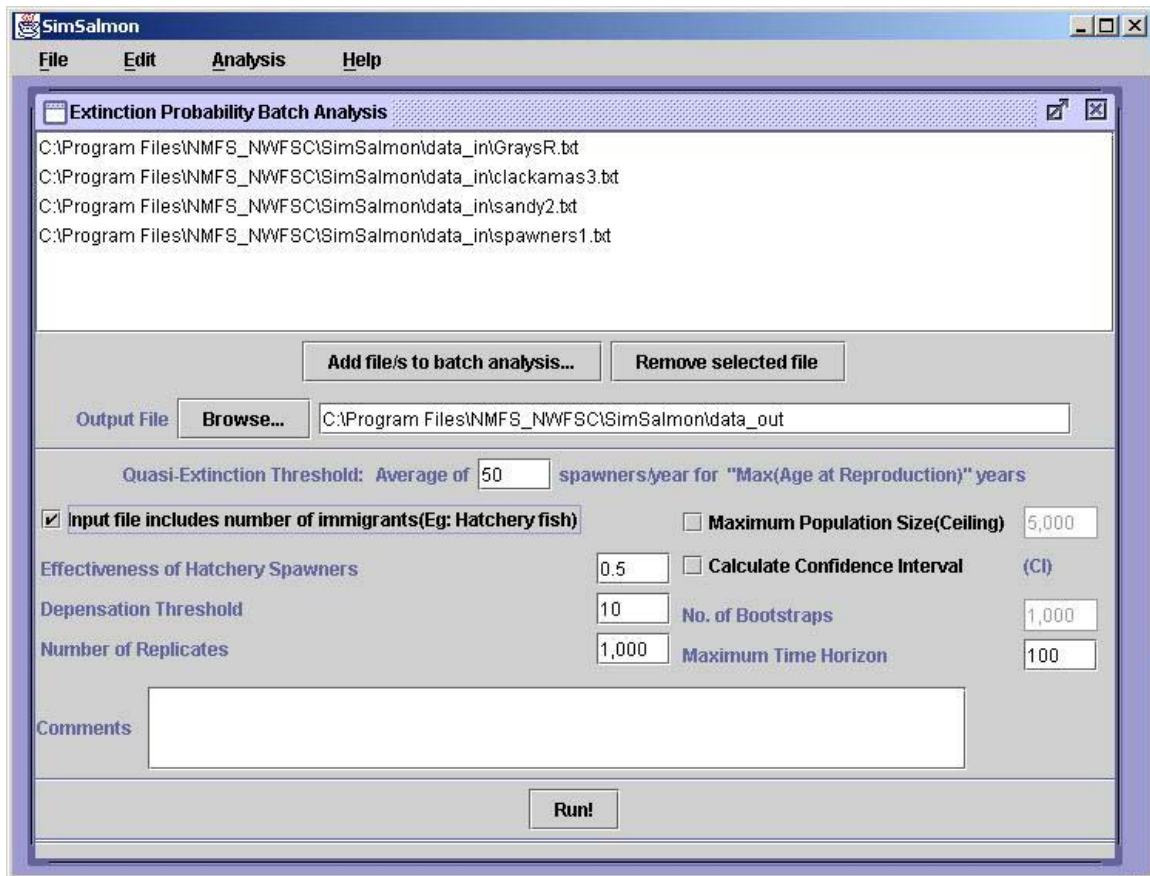


Figure 11: Batch Extinction Calculator input screen.

Sensitivity Analysis Sub-Module

The sensitivity analysis input screens are similar to the extinction calculator input screen. However, only a single time horizon can be examined, trajectories cannot be written to a file and a number of parameter values are evaluated for a single click of the Run! button. The program loops through a series of parameter values beginning with “Start <parm. Name>”, incrementing by “Increment <param. Name>” and ending with “Stop <param. Name>”. In the initial size sensitivity analysis, all maxAge initial years are set to the same value. The output file is pretty self-explanatory.

SimSalmon

File Edit Analysis Help

Population Ceiling

Maximum Age at Reproduction: 7

Age Distribution of Spawners

Age1	Age2	Age3	Age4	Age5	Age6	Age7
0.0	0.0	0.0	0.0	0.0	0.0	0.0

Initial Population Size

Year1	Year2	Year3	Year4	Year5	Year6	Year7
0	0	0	0	0	0	0

Quasi-Extinction Threshold: Average of 50 spawners/year for 7 years

Calculation Type: Spawner:Spawner

Average Spawner:Spawner: 1

Variance Spawner:Spawner: 1

☐ Maximum Spawner:Spawner: 3

Number of Replicates: 1,000

Depensation Threshold: 10

Time Horizon: 100

☐ Calculate Confidence Interval (CI)

No. of Bootstraps:

Sample Size: 0

Ceiling Start: 100

Ceiling Increment: 100

Ceiling Stop: 1,000

Comments:

Output File: Browse...

Run!

Figure 12: “Population Ceiling” sensitivity analysis input screen.

Patterns in Spawners/Spawner Sub-Module

TEMPORAL AUTOCORRELATION

In “Temporal Autocorrelation”, the user specifies the maximum number of lags and the autocorrelation values for $\text{mean}(\ln(r/s))$. The technique used to simulate a normally distributed, temporally autocorrelated pattern relies on a Cholesky decomposition of the covariance matrix and the covariance matrix must be symmetrical definite positive. The simulation technique involves generation of a sequence of $\ln(r/s)$ values with the proper autocorrelation structure using the following equation.

$$y = Lw + m,$$

where y is the simulated vector of $\ln(r/s)$ values (the size of the vector is the maximum time horizon), L is the lower triangular matrix of the Cholesky decomposition of the variance-covariance matrix (size = time horizon X time horizon), w is a vector of independent random standard normal deviates, and m is a vector with every element equal to $\text{mean}(\ln(r/s))$. The variance-covariance matrix is created base on $\text{var}(\ln(r/s))$ and the specified autocorrelation values. Table 15 shows an example covariance matrix. The simulation approach is described by Davis (1987, Mathematical Geology 19(2):91-98). A new y vector is generated for each trajectory, and $y(t)$ is used as the $\ln(r/s)$ value at time t . The input window is shown in Figure 12. The output looks like the other extinction risk calculations.

Table 15: Example covariance matrix for simulating temporal autocorrelation. The elements of the matrix are the expected covariance between values of $\ln(r/s)$ at different time periods in the trajectory. In this example the overall variance is 0.5 and the autocorrelations are $\text{lag1} = 0.6$, $\text{lag2} = 0.3$. At lags greater than 2, the autocorrelation is zero. The covariance is the autocorrelation times the variance. The lower triangular matrix of a covariance matrix such as this one would be entered as L in the simulation equation above. This matrix is generated by the program and only the variance and autocorrelation values need to be specified by the user.

Time	1	2	3	4	5	6	...	100
1	0.5	0.3	0.15	0	0	0	...	0
2	0.3	0.5	0.3	0.15	0	0	...	0
3	0.15	0.3	0.5	0.3	0.15	0	...	0
4	0	0.15	0.3	0.5	0.3	0.15	...	0
5	0	0	0.15	0.3	0.5	0.3	...	0
6	0	0	0	0.15	0.3	0.5	...	0
...
100	0	0	0	0	0	0	...	0.5

The screenshot shows the 'Temporal Autocorrelation' window in the SimSalmon software. The window has a menu bar with 'File', 'Edit', 'Analysis', and 'Help'. The main area contains several input sections:

- Maximum Age at Reproduction:** A text box with the value '5'.
- Age Distribution of Spawners:** A table with columns 'Age1' through 'Age5' and rows for 'Age1' through 'Age5'. The values are mostly 0.0.
- Initial Population Size:** A table with columns 'Year1' through 'Year5' and rows for 'Year1' through 'Year5'. The values are mostly 0.
- Quasi-Extinction Threshold:** A text box with the value '50' and a label 'Average of 50 spawners/year for 5 years'.
- Ln(Spawner:Spawner):** A text box with the value '0.7'.
- Number of Replicates:** A text box with the value '1,000'.
- Dispensation Threshold:** A text box with the value '10'.
- Maximum Time Horizon:** A text box with the value '100'.
- Maximum Autocorrelated Lag:** A text box with the value '5'.
- Maximum Population Size(Ceiling):** A text box with the value '5,000'.
- Lag Autocorrelations:** A table with columns 'Lag1' through 'Lag5' and rows for 'Lag1' through 'Lag5'. The values are mostly 0.0.
- Confidence Interval(CI):** A checkbox labeled 'Confidence Interval(CI)'.
- No. of Bootstraps:** A text box.
- Sample Size:** A text box.
- Comments:** A large text area.
- Write Trajectories to File:** A checkbox.
- Output File:** A text box with a 'Browse...' button.
- Run!:** A button.

Figure 12: Input window for temporal autocorrelation extinction risk estimator.

TRENDS IN SPAWNERS PER SPAWNER

This analysis allows estimation of extinction risk when there is a linear trend in $\text{mean}(r/s)$. The user specifies the initial $\text{mean}(r/s)$ and $\text{mean}(r/s)$ is changed during the

course of each trajectory as a linear function of time based on the slope input. If mean r/s is < 0 , the population is considered extinct.

Appendix A: Correction for hatcheries and harvest

Slope Method of Parameter Estimation Corrections for Hatchery

This is slightly modified presentation of the approach used to by the CRI (Holmes 2000) to correct for the input of hatchery spawners when using the slope method to estimated extinction risk parameters. First, consider a simple population model of stochastic exponential growth for a semelparous annual . Let W_t be the number of individuals at time t and let r be the stochastic per capita growth rate. The parameters we want to estimate are μ and σ^2 , the parameters describing the distribution of r .

$$W_{t+1} = W_t e^r$$

$$W_{t+2} = W_{t+1} e^r$$

$$W_{t+2} = (W_t e^r) e^r = W_t e^{2r}$$

$$\ln\left(\frac{W_{t+1}}{W_t}\right) = r \approx \text{Norm}(\mu, \sigma^2)$$

$$\ln\left(\frac{W_{t+2}}{W_t}\right) = 2r \approx \text{Norm}(2\mu, 2\sigma^2)$$

$$\ln\left(\frac{W_{t+\tau}}{W_t}\right) = \tau r \approx \text{Norm}(\tau\mu, \tau\sigma^2)$$

$$\mu = \text{slope of } \text{mean}\left(\ln\left(\frac{W_{t+\tau}}{W_t}\right)\right) \text{ vs. } \tau$$

$$\sigma^2 = \text{slope of } \text{var}\left(\ln\left(\frac{W_{t+\tau}}{W_t}\right)\right) \text{ vs. } \tau$$

Now, consider a population model of stochastic exponential growth where some fraction of the breeding population every generation is composed of first generation hatchery produced fish. The population is annual semelparous as before. Let W_t be the number of natural origin (“wild”) spawners at time t , H_t be the number of hatchery origin spawners at time t . Let θ_t be the fraction of hatchery origin spawners reproducing in the wild at time t and let r be as before. Let ψ_t be the ratio of total spawners to natural origin spawners at time t .

$$\theta_t = \frac{H_t}{W_t + H_t}$$

$$\psi_t = \frac{1}{1 - \theta_t}$$

$$W_{t+1} = \psi_t W_t e^r$$

$$W_{t+2} = \psi_{t+1} W_{t+1} e^r$$

$$W_{t+2} = \psi_{t+1} (\psi_t W_t e^r) e^r = \psi_t \psi_{t+1} W_t e^{2r}$$

$$\ln\left(\frac{W_{t+1}}{\psi_t W_t}\right) = \ln\left(\frac{W_{t+1}}{W_t}\right) - \ln(\psi_t) = r \approx \text{Norm}(\mu, \sigma^2)$$

$$\ln\left(\frac{W_{t+2}}{\psi_t \psi_{t+1} W_t}\right) = \ln\left(\frac{W_{t+2}}{W_t}\right) - \ln(\psi_t \psi_{t+1}) = 2r \approx \text{Norm}(2\mu, 2\sigma^2)$$

$$\ln\left(\frac{W_{t+\tau}}{W_t}\right) - \ln\left(\prod_{i=0}^{\tau-1} \psi_{t+i}\right) = \tau r \approx \text{Norm}(\tau\mu, \tau\sigma^2)$$

$$\mu = \text{slope of } \text{mean}\left(\ln\left(\frac{W_{t+\tau}}{W_t}\right) - \ln\left(\prod_{i=0}^{\tau-1} \psi_{t+i}\right)\right) \text{ vs. } \tau$$

$$\sigma^2 = \text{slope of } \text{var}\left(\ln\left(\frac{W_{t+\tau}}{W_t}\right) - \ln\left(\prod_{i=0}^{\tau-1} \psi_{t+i}\right)\right) \text{ vs. } \tau$$

To account for the overlapping semelparous age structure observed in salmon, the above equations must be modified as follows.

$$\mu = \text{slope of } \text{mean} \left(\ln \left(\frac{W_{t+\tau}}{W_t} \right) - \frac{1}{T} \ln \left(\prod_{i=0}^{\tau-1} \psi_{t+i} \right) \right) \text{ vs. } \tau$$

$$\sigma^2 = \text{slope of } \text{var} \left(\ln \left(\frac{W_{t+\tau}}{W_t} \right) - \frac{1}{T} \ln \left(\prod_{i=0}^{\tau-1} \psi_{t+i} \right) \right) \text{ vs. } \tau$$

, where T is the mean age at reproduction. These equations are derived via an age-structured model as follows.

$$\mathbf{W}_{t+1} = \mathbf{A}_t \times (\mathbf{N}_t + \mathbf{K}_t),$$

where \mathbf{N}_t is a vector of natural origin spawners at time t , \mathbf{A}_t is the underlying stochastic matrix for which we are trying to estimate μ and σ^2 , and \mathbf{K}_t is a vector of hatchery origin input at time t and takes the form:

$$\mathbf{K}_t = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dots \\ S_{h,t} \end{bmatrix}$$

where $S_{h,t}$ is the number of hatchery origin spawners at time t . The vector \mathbf{K}_t can be defined in terms of \mathbf{W}_t as

$$\mathbf{K}_t = \mathbf{J}_t \times \mathbf{W}_t$$

where

$$\mathbf{J}_t = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & \psi_t \theta_t \end{bmatrix}$$

Combining equations yields

$$\mathbf{W}_{t+1} = \mathbf{A}_t \times (\mathbf{W}_t + \mathbf{J}_t \times \mathbf{W}_t) = \mathbf{A}_t \times (\mathbf{I} + \mathbf{J}_t) \times \mathbf{W}_t$$

This can be rewritten as

$$\mathbf{W}_{t+1} = \mathbf{B}_t \times \mathbf{W}_t$$

where

$$\mathbf{B}_t = \mathbf{A}_t \times (\mathbf{I} + \mathbf{J}_t) = \mathbf{A}_t \times \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 + \psi_t \theta_t \end{bmatrix} = \mathbf{A}_t + \begin{bmatrix} 0 & 0 & 0 & \dots & F * \psi_t \theta_t \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

where F is the mean fecundity rate. If two projection matrices differ only in the fecundity term, the ratio of the net reproductive values is equivalent to the ratio of the fecundity terms (Caswell 1989). Since \mathbf{A}_t and \mathbf{B}_t differ only in the fecundity term, the net reproductive value of \mathbf{B}_t is related to the net reproductive value of \mathbf{A}_t as

Let R_{0,A_t} = net reproductive value of \mathbf{A}_t

Let R_{0,B_t} = net reproductive value of \mathbf{B}_t

$$R_{0,B_t} = R_{0,A_t} \times \frac{F(1 + \psi_t \theta_t)}{F} = R_{0,A_t} \psi_t$$

$$R_{0,A_t} = R_{0,B_t} \left(\frac{1}{\psi_t} \right)$$

The relationship between the net reproductive rate and μ is (Caswell 1989)

$$\lambda^T = R_0$$

$$T(\ln(\lambda)) = T\mu = \ln(R_0)$$

We can now solve for μ_{A_t}

$$\mu_{B_t} = \frac{1}{T} \ln(R_{0,B_t}) = \ln\left(\frac{W_{t+1}}{W_t}\right)$$

$$\mu_{A_t} = \frac{1}{T} \ln(R_{0,A_t}) = \frac{1}{T} \ln\left(R_{0,B_t} \left(\frac{1}{\psi_t}\right)\right) = \mu_{B_t} - \frac{1}{T} \ln(\psi_t) = \ln\left(\frac{W_{t+1}}{W_t}\right) - \frac{1}{T} \ln(\psi_t)$$

We end up with the original equations

$$\hat{\mu} = \text{mean}\left(\ln\left(\frac{W_{t+1}}{W_t}\right) - \frac{1}{T} \ln(\psi_t)\right)$$

$$\hat{\sigma}^2 = \text{var}\left(\ln\left(\frac{W_{t+1}}{W_t}\right) - \frac{1}{T} \ln(\psi_t)\right)$$

Which is extended as in the annual case to estimate via the slope method

$$\hat{\mu} = \text{slope of } \text{mean}\left(\ln\left(\frac{W_{t+\tau}}{W_t}\right) - \frac{1}{T} \ln\left(\prod_{i=0}^{\tau-1} \psi_{t+i}\right)\right) \text{ vs. } \tau$$

$$\hat{\sigma}^2 = \text{slope of } \text{var}\left(\ln\left(\frac{W_{t+\tau}}{W_t}\right) - \frac{1}{T} \ln\left(\prod_{i=0}^{\tau-1} \psi_{t+i}\right)\right) \text{ vs. } \tau$$

The total population size, W_t is estimated by the weighted running sum

$$W_t = S_t + \sum_{i=1}^{\max age-1} w_i S_{t-i}$$

where S_t is the number of natural origin spawners in year t and $w_i = 1 - \sum_{j=1}^i fractionAge_j$.

The fraction of hatchery origin spawners is also calculated based on the weighted running sum.

Slope Method of Parameter Estimation Corrections for Harvest

This formulation is analogous to the approach of correcting for the input of hatchery spawners when using the slope method to estimated extinction risk parameters (CRI 2000). First, consider a simple discrete population model of stochastic exponential growth. Let N_t be the number of individuals at time t and let r be the stochastic per capita growth rate. The parameters we want to estimate are μ and σ^2 , the parameters describing the distribution of r .

$$N_{t+1} = N_t e^r$$

$$N_{t+2} = N_{t+1} e^r$$

$$N_{t+2} = (N_t e^r) e^r = N_t e^{2r}$$

$$\ln\left(\frac{N_{t+1}}{N_t}\right) = r \approx \text{Norm}(\mu, \sigma^2)$$

$$\ln\left(\frac{N_{t+2}}{N_t}\right) = 2r \approx \text{Norm}(2\mu, 2\sigma^2)$$

$$\ln\left(\frac{N_{t+\tau}}{N_t}\right) = \tau r \approx \text{Norm}(\tau\mu, \tau\sigma^2)$$

$$\mu = \text{slope of } \text{mean}\left(\ln\left(\frac{N_{t+\tau}}{N_t}\right)\right) \text{ vs. } \tau$$

$$\sigma^2 = \text{slope of } \text{var}\left(\ln\left(\frac{N_{t+\tau}}{N_t}\right)\right) \text{ vs. } \tau$$

Now, consider a discrete population model of stochastic exponential growth where some fraction of the population is harvested every generation. Let N_t be the number of individuals that would have returned at time t if there had been no harvest, let C_t be the catch at time t and W_t be the number of spawners at time t . Let ϕ_t be the harvest rate at time t and let r be as before. Let φ_t be the fraction not harvested at time t .

$$N_t = C_t + W_t$$

$$\phi_t = \frac{C_t}{C_t + W_t}$$

$$\varphi_t = 1 - \phi_t$$

$$N_{t+1} = \varphi_t N_t e^r$$

$$N_{t+2} = \varphi_{t+1} N_{t+1} e^r$$

$$N_{t+2} = \varphi_{t+1} (\varphi_t N_t e^r) e^r = \varphi_t \varphi_{t+1} N_t e^{2r}$$

$$\ln\left(\frac{N_{t+1}}{\varphi_t N_t}\right) = \ln\left(\frac{N_{t+1}}{N_t}\right) - \ln(\varphi_t) = r \approx \text{Norm}(\mu, \sigma^2)$$

$$\ln\left(\frac{N_{t+2}}{\varphi_t \varphi_{t+1} N_t}\right) = \ln\left(\frac{N_{t+2}}{N_t}\right) - \ln(\varphi_t \varphi_{t+1}) = 2r \approx \text{Norm}(2\mu, 2\sigma^2)$$

$$\ln\left(\frac{N_{t+\tau}}{N_t}\right) - \ln\left(\prod_{i=0}^{\tau-1} \varphi_{t+i}\right) = \tau r \approx \text{Norm}(\tau\mu, \tau\sigma^2)$$

$$\mu = \text{slope of } \text{mean}\left(\ln\left(\frac{N_{t+\tau}}{N_t}\right) - \ln\left(\prod_{i=0}^{\tau-1} \varphi_{t+i}\right)\right) \text{ vs. } \tau$$

$$\sigma^2 = \text{slope of } \text{var}\left(\ln\left(\frac{N_{t+\tau}}{N_t}\right) - \ln\left(\prod_{i=0}^{\tau-1} \varphi_{t+i}\right)\right) \text{ vs. } \tau$$

To account for the overlapping semelparous age structure observed in salmon, the above equations must be modified as follows.

$$\mu = \text{slope of } \text{mean} \left(\ln \left(\frac{N_{t+\tau}}{N_t} \right) - \frac{1}{T} \ln \left(\prod_{i=0}^{\tau-1} \phi_{t+i} \right) \right) \text{ vs. } \tau$$

$$\sigma^2 = \text{slope of } \text{var} \left(\ln \left(\frac{N_{t+\tau}}{N_t} \right) - \frac{1}{T} \ln \left(\prod_{i=0}^{\tau-1} \phi_{t+i} \right) \right) \text{ vs. } \tau$$

, where T is the mean age at reproduction. These equations are derived via an age-structured model as follows.

$$\mathbf{N}_{t+1} = \mathbf{A}_t \times (\mathbf{N}_t - \mathbf{G}_t),$$

where \mathbf{N}_t is a vector of the fish that would be alive at time t if there had been no harvest, \mathbf{A}_t is the underlying stochastic matrix for which we are trying to estimate μ and σ^2 , and \mathbf{G}_t is a vector of fish removed by the harvest at time t and takes the form:

$$\mathbf{G}_t = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dots \\ S_{c,t} \end{bmatrix}$$

where $S_{c,t}$ is the number of spawners that were removed by harvest at time t . The vector \mathbf{G}_t can be defined in terms of \mathbf{N}_t as

$$\mathbf{G}_t = \mathbf{Q}_t \times \mathbf{N}_t$$

where

$$\mathbf{Q}_t = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & \phi_t \end{bmatrix}$$

with ϕ_t redefined as the number of fish that did not returned to spawn at time t because of harvest divided by the total number of spawners that would have returned at time t if there had been no harvest. Note that this is not an attempt to partition out age specific mortalities, but simple looks at how many fish did not come back because of harvest.

Combining equations yields

$$\mathbf{N}_{t+1} = \mathbf{A}_t \times (\mathbf{N}_t - \mathbf{Q}_t \times \mathbf{N}_t) = \mathbf{A}_t \times (\mathbf{I} - \mathbf{Q}_t) \times \mathbf{N}_t$$

This can be rewritten as

$$\mathbf{N}_{t+1} = \mathbf{D}_t \times \mathbf{N}_t$$

where

$$\mathbf{D}_t = \mathbf{A}_t \times (\mathbf{I} - \mathbf{Q}_t) = \mathbf{A}_t \times \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 - \phi_t \end{bmatrix} = \mathbf{A}_t + \begin{bmatrix} 0 & 0 & 0 & \dots & -F^* \phi_t \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

where F is the mean fecundity rate. If two projection matrices differ only in the fecundity term, the ratio of the net reproductive values is equivalent to the ratio of the fecundity terms (Caswell 1989). Since \mathbf{A}_t and \mathbf{D}_t differ only in the fecundity term, the net reproductive value of \mathbf{D}_t is related to the net reproductive value of \mathbf{A}_t as

Let R_{0,A_t} = net reproductive value of \mathbf{A}_t

Let R_{0,D_t} = net reproductive value of \mathbf{D}_t

$$R_{0,D_t} = R_{0,A_t} \times \frac{F(1 - \phi_t)}{F} = R_{0,A_t} \phi_t$$

$$R_{0,A_t} = R_{0,D_t} \left(\frac{1}{\phi_t} \right)$$

where $\phi_t = 1 - \phi_t$ as above.

The relationship between the net reproductive rate and μ is (Caswell 1989)

$$\lambda^T = R_0$$

$$T(\ln(\lambda)) = T\mu = \ln(R_0)$$

We can now solve for μ_{A_t}

$$\mu_{D_t} = \frac{1}{T} \ln(R_{0,D_t}) = \ln\left(\frac{N_{t+1}}{N_t}\right)$$

$$\mu_{A_t} = \frac{1}{T} \ln(R_{0,A_t}) = \frac{1}{T} \ln\left(R_{0,D_t} \left(\frac{1}{\phi_t}\right)\right) = \mu_{D_t} - \frac{1}{T} \ln(\phi_t) = \ln\left(\frac{N_{t+1}}{N_t}\right) - \frac{1}{T} \ln(\phi_t)$$

We end up with the original equations

$$\hat{\mu} = \text{mean}\left(\ln\left(\frac{N_{t+1}}{N_t}\right) - \frac{1}{T} \ln(\phi_t)\right)$$

$$\hat{\sigma}^2 = \text{var}\left(\ln\left(\frac{N_{t+1}}{N_t}\right) - \frac{1}{T} \ln(\phi_t)\right)$$

Which is extended as in the discrete case to estimate via the slope method

$$\hat{\mu} = \text{slope of } \text{mean} \left(\ln \left(\frac{N_{t+\tau}}{N_t} \right) - \frac{1}{T} \ln \left(\prod_{i=0}^{\tau-1} \varphi_{t+i} \right) \right) \text{ vs. } \tau$$

$$\hat{\sigma}^2 = \text{slope of } \text{var} \left(\ln \left(\frac{N_{t+\tau}}{N_t} \right) - \frac{1}{T} \ln \left(\prod_{i=0}^{\tau-1} \varphi_{t+i} \right) \right) \text{ vs. } \tau$$

The total population size, N_t is estimated by the weighted running sum

$$N_t = S_t + \sum_{i=1}^{\max \text{ age}-1} w_i S_{t-i}$$

where S_t is the number of spawners that would have returned in year t if there had been no harvest and $w_i = 1 - \sum_{j=1}^i \text{fractionAge}_j$.

The fraction harvested is also calculated based on the weighted running sum.

Slope Method of Parameter Estimation Corrections for Hatchery and Harvest

This is derived by combining the hatchery and harvest analyses based on the equation

$$W_{t+1} + C_{t+1} = N_{t+1} = \varphi_t \psi_t N_t e^r = \left(\frac{W_t}{C_t + W_t} \right) \left(\frac{W_t + H_t}{W_t} \right) (C_t + W_t) e^r = (W_t + H_t) e^r$$

$$r = \ln \left(\frac{W_{t+1} + C_{t+1}}{W_t + H_t} \right) \approx \text{Norm}(\mu, \sigma^2)$$

Which is extended as in the annual case to estimate via the slope method

$$\hat{\mu} = \text{slope of } \text{mean} \left(\ln \left(\frac{N_{t+\tau}}{N_t} \right) - \frac{1}{T} \ln \left(\prod_{i=0}^{\tau-1} \varphi_{t+i} \psi_{t+i} \right) \right) \text{ vs. } \tau$$

$$\hat{\sigma}^2 = \text{slope of } \text{var} \left(\ln \left(\frac{N_{t+\tau}}{N_t} \right) - \frac{1}{T} \ln \left(\prod_{i=0}^{\tau-1} \varphi_{t+i} \psi_{t+i} \right) \right) \text{ vs. } \tau$$

The total population size, N_t is estimated by the weighted running sum

$$N_t = S_t + \sum_{i=1}^{\max \text{ age}-1} w_i S_{t-i}$$

where S_t is the number of spawners that would have returned in year t if there had been no harvest and $w_i = 1 - \sum_{j=1}^i \text{fractionAge}_j$.

The fraction of hatchery origin spawners is also calculated based on the weighted running sum.

Appendix B: Variance degrees of freedom

SimSalmon uses the slope method (Holmes 2001) for estimating the variance in $\ln(N_{t+1}/N_t)$. The slope method can greatly reduce the bias in the variance estimate created by measurement error. However, the reduction in bias comes at the cost of decreased precision. Holmes and Fagan (unpublished manuscript) investigate the sampling distribution of the mean and variance of $\ln(N_{t+1}/N_t)$. Using analytical and simulation approaches, they demonstrate that the ratio of the variance estimate to the true variance is approximated by the typical X^2 distribution and the mean is approximated by the Student's t distribution. However the number of degrees of freedom for the X^2 and t distributions are greatly reduced from the typical value calculated as the number of observations minus one. In Table B1, Holmes and Fagan list approximate degrees of freedom for the slope method variance estimate as a function of the number of years of the time series and the length of the running sum. The degrees of freedom increases approximately linearly with increasing years of data and SimSalmon uses linear equations to calculate the degrees of freedom (Figures B1 and B2). The equation used depends on whether weighted or unweighted running sums are used. For the weighted running sum, the number of estimates of $\ln(N_{t+1}/N_t)$ is used as the sample size and the equation was determined as shown in Figure B1. The unweighted sum uses the number of years of spawner counts and the length of the running sum as inputs (Figure B2).

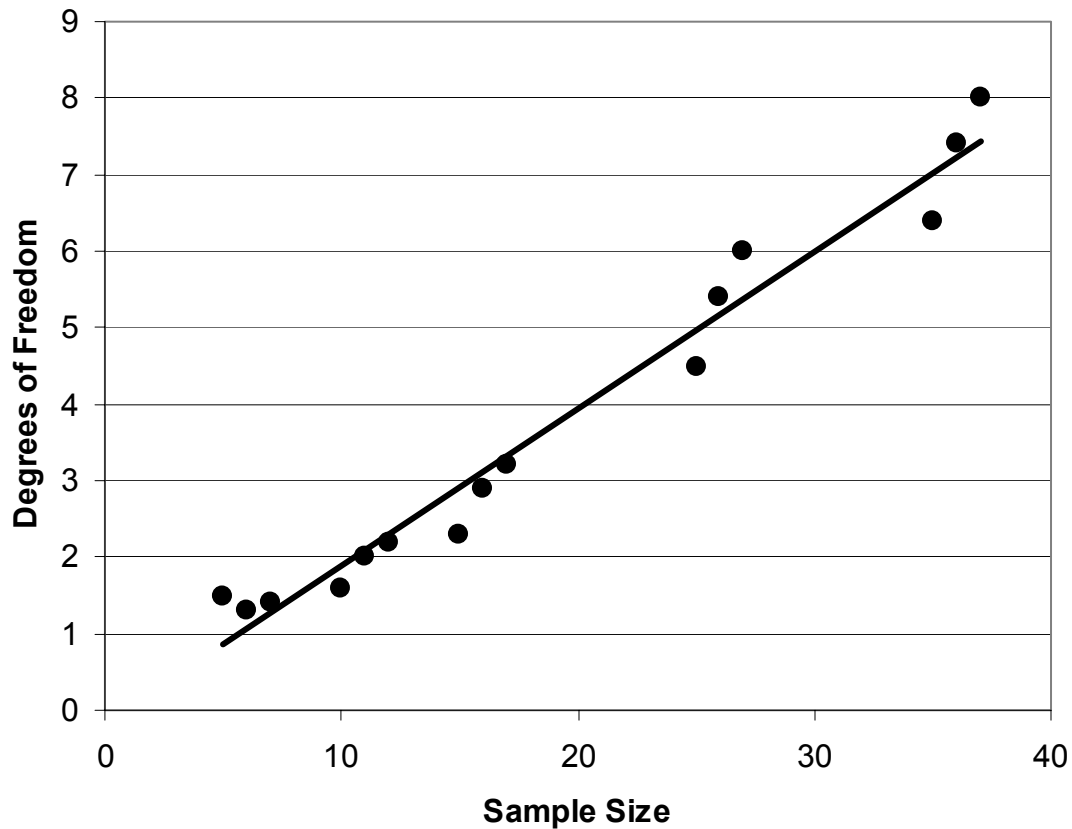


Figure B1: Relationship between sample size and the degrees of freedom associated with the distribution of the slope method variance estimate. Figure data is taken from Holmes and Fagan (unpublished manuscript), Table B1. To produce this figure from the Holmes and Fagan table, sample size is calculated by subtracting the length of the running sum, L , from the number of years. This loosely approximates the number of estimates of $\ln(N_{t+1}/N_t)$, which is the value for sample size used in SimSalmon when the weighted running sum option is selected. The degrees of freedom are calculated by assuming a non-process variance to process variance ratio of one (see Holmes and Fagan). The equation for the linear fit that is used by SimSalmon is $df = 0.21 * \text{sampleSize} - 0.17$.

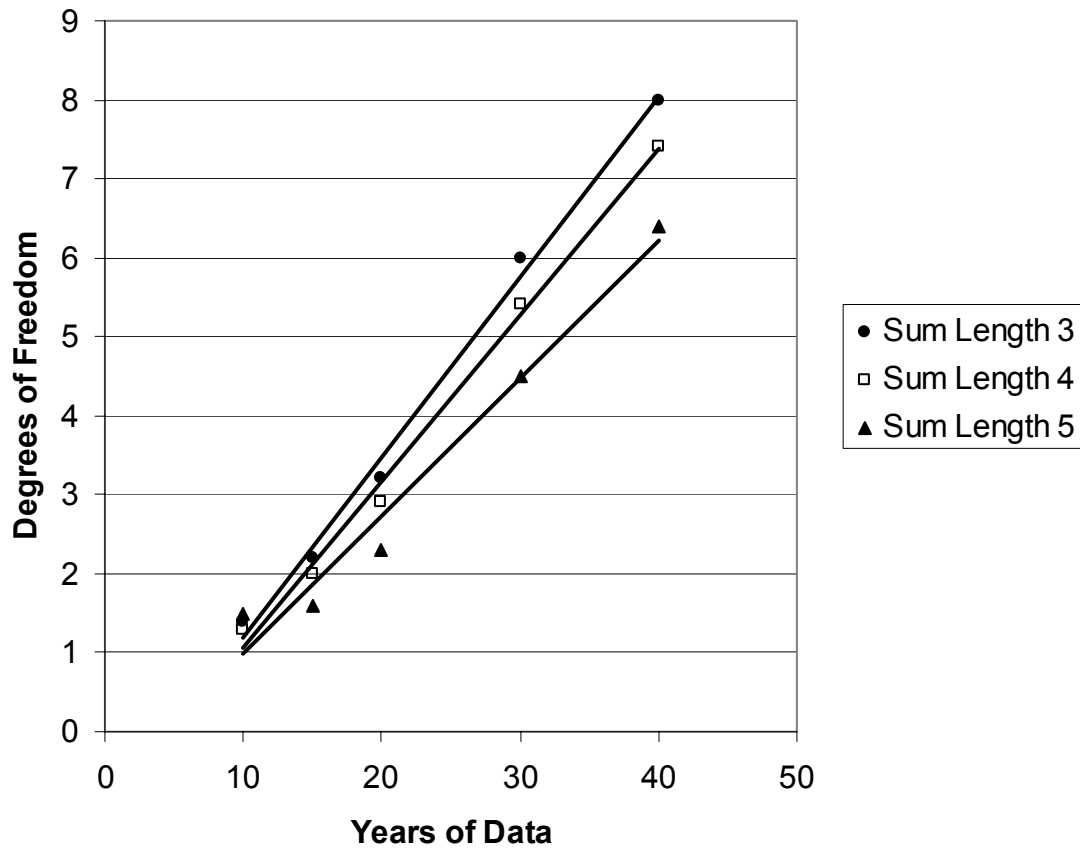


Figure B2: Relationship between sample size and the degrees of freedom associated with the distribution of the slope method variance estimate. Figure data is taken from Holmes and Fagan (unpublished manuscript), Table B1. These are the relationships used when the unweighted running sum option is selected in SimSalmon. The degrees of freedom are calculated by assuming a non-process variance to process variance ratio of one (see Holmes and Fagan). The equation for the linear fit that is used by SimSalmon with a running sum of length three is $df = 0.23 * nYears - 1.10$. For a running sum of length four the equation is $df = 0.21 * nYears - 1.04$. For a running sum of length five the equation is $df = 0.17 * nYears - 0.75$.